

Investigating the impact of chromosomal rearrangements on the evolutionary history of rodents leading to *Mus musculus*

A thesis to the University of Kent for the degree of

M.Sc. by Research in Computational Biology

2019

Emma Leach

School of Biosciences

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

Acknowledgements

The greatest portion of my thanks has to go to my supervisor Marta. Marta has been wonderful support to me during this year, despite being new to the university and all of its processes and systems. She has taught me many new skills, pushed me when I needed pushing, helped me when I got stuck, and laughed at me (nicely) when I have run myself down an overcomplicated rabbit hole. She has also provided me with a great deal of advice regarding my options after this research, which I am deeply grateful for as it is an area in which I am chronically indecisive.

There are a number of scripts used or adapted during the process of this research. Thanks for these are extended to Marta, Jaebum Kim, and Joana Damas.

Thank you to the various members of the Wass-Michaelis lab who have made me feel very welcome, even if they have struck me off the friendship plate now.

A huge thanks also to my two housemates Ross and Elijah, who have both provided me with lessons in programming and 2am McDonalds trips. Elijah in particular has helped me battle imposter syndrome throughout this year, long after ceasing to be housemates, and has kindly hosted the mySynteny portal for me.

I could not class myself as an animal lover without thanking the animals in my life who have provided me more emotional comfort than any human could. So, thank you Skittles, who has trampled over my keyboard so many times it is a miracle I have any valid data at all. And thank you to Bruno, who has ensured that I haven't slept through the write up stage of this process by constantly stealing my bed.

Contents

Declaration *2*

Acknowledgements *3*

Contents *4*

List of Figures *7*

List of Tables *10*

Abbreviations *11*

Abstract *13*

Introduction *14*

Speciation and Genome Evolution **14**

Types of Genomic Rearrangements 16

Models for Chromosomal Speciation 21

Models for Genome Evolution 23

Methods for Predicting Ancestral Karyotypes **26**

Cytogenetic Methods 26

Computational Methods 27

Order Rodentia **30**

Project Aims **33**

Materials and Methods *34*

Reconstruction of Rodentia Ancestors **34**

Genomic Data 34

Phylogenetic Tree Construction 35

Pairwise Alignments 35
Reconstruction of Rodentia Ancestors 36
Identification of Chromosome Rearrangements 36
Gene Expression Analysis 37
RNA-Seq Data 37
RNA-Seq Alignment and Gene Counts 37
Filtering for Orthologues 38
Correlation of Gene Expression 38
Gene Expression in Rearrangements 38
Gene Ontology (GO) Enrichment Analysis 39
Results and Discussion 40
Reconstruction of Rodentia Ancestors 40
Genome Selection and Alignment 40
Phylogenetic Trees 44
Reconstructing Ancestral Predicted Chromosome Fragments 47
Comparison to Cytogenetic Studies – Muridae 53
Comparison to Cytogenetic Studies – Eumuroidea 56
Comparison to Cytogenetic Studies – Muroidea 58
Comparison to Cytogenetic Studies - Rodentia 61
Identification of Chromosomal Rearrangements 64
Rates of Chromosomal Rearrangements 65
Discussion 67
Gene Expression Analysis 71
RNA-Seq Alignment and Gene Counts 71
Filtering for Orthologous Genes 74
Principal Component Analysis 74
Gene Expression Correlation 76

Gene Expression in Rearrangements 77

Orthologues Absent from Muridae Reconstruction 77

Rearrangements in Inversions - Liver 79

Rearrangements in Inversions - Testes 82

Gene Ontology in Inversions 84

Discussion 86

Conclusion 89

References 90

Appendix 111

List of Figures

Figure 1 - Chromosomal inversion (image modified from 'Chromosomenmutationen' - Wikimedia commons) . 17

Figure 2 - Meiotic products resulting from a paracentric inversion heterozygote (left) and a paracentric inversion heterozygote (right) [17] 18

Figure 3 - Chromosomal translocation (image modified from 'Chromosomenmutationen' - Wikimedia commons) 19

Figure 4 - Chromosomal deletion (image modified from 'Chromosomenmutationen' - Wikimedia commons)... 20

Figure 5 - Chromosomal duplication and insertion (image modified from 'Chromosomenmutationen' - Wikimedia commons) 21

Figure 6 - Summary of the hybrid sterility model and recombination suppression model of chromosomal speciation [18] 22

Figure 7 – Chromosome painting carried out on human chromosomes using chromosome-specific paint probes derived from gibbon chromosomes [49] 27

Figure 8 - Evolutionary tree of the order Rodentia, showing many of the major families [91] 32

Figure 9 - Heatmap of the coverage of rodent net files as a fraction when compared to Mus musculus..... 43

Figure 10 - Evolutionary trees for the Order Rodentia [A] where the squirrel related lineage (red) is the closest relation to the mouse related clade. [B] where Ctenohystricia is the closest relation to the mouse related clade, and the squirrel related lineage (red) is the 'root' of the rodent evolutionary tree. 46

Figure 11 - mySynteny view of syntenic blocks shared between Muridae ancestor APCF4 and various APCFs in the Ctenohystricia ancestor 51

Figure 12 - Mus musculus chromosome 15 in relation to the APCFs of each of the predicted ancestors in the Evolution Highway format. Blue and pink blocks represent syntenic fragments in "+" (blue) and "-" (pink) orientation. The number in the block represents the APCF reference number. 52

Figure 13 - Mus musculus chromosome 17 in relation to the APCFs of each of the predicted ancestors in the Evolution Highway format. Blue and pink blocks represent syntenic fragments in "+" (blue) and "-" (pink) orientation. The number in the block represents the APCF reference number. 53

Figure 14 - Ancestral karyotype for the ancestor of Muridae. Different colours correspond to separate mouse chromosomes [126] 53

Figure 15 - Ideogram of APCFs for Muridae ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from mouse chromosomes..... 54

Figure 16 - Ancestral Eumuroidea karyotype. The homologies of mouse (MMU) chromosomes are shown to the left of the ideogram, and the homologies of golden hamster (MAU) chromosomes are shown to the right of the ideogram. Number in bold represent possible homology with human chromosome fragments [128] 56

Figure 17 - Ideogram of APCFs for Eumuroidea ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from mouse chromosomes..... 57

Figure 18 - Ancestral karyotype for the ancestor of Muroidea. Different colours correspond to separate mouse chromosomes [126] 58

Figure 19 - Ancestral Muroidea karyotype. The homologies of mouse (MMU) chromosomes are shown to the left of the ideogram, and the homologies of golden hamster (MAU) chromosomes are shown to the right of the ideogram..... 59

Figure 20 - Ideogram of APCFs for Muroidea ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from mouse chromosomes..... 60

Figure 21 – [A] Ancestral karyotype of the Rodentia ancestor from cross-species chromosome painting results. Different numbers correspond to homologies of individual conserved segments in human chromosomes (HSA) [43] [B] Ancestral karyotype of the Rodentia ancestor from cross-species chromosome painting results from comparative squirrel genome maps, and on alignments of mouse, rat and human genome sequences. Different numbers correspond to homologies of individual conserved segments in human chromosomes (HSA) [152].... 62

Figure 22 - Ideogram of APCFs for Rodentia ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from human chromosomes..... 63

Figure 23 - Phylogenetic tree of rodent species showing rate of chromosomal inversions between reconstructed ancestors. Numbered nodes represent the following ancestors: 1 – Muridae, 2 – Eumuroidea, 3 – Muroidea, 4 – Myodonta, 5 – Mouse lineage, 6 – Mouse lineage + Ctenohystricia, 7 – Rodentia 67

Figure 24 - Gene counts for Mus musculus liver RNA-Seq data under different STAR and HTSeq parameters. Samples denoted with '_1' were treated as single end, with a non strand specific protocol. Samples denoted

with '_2' were treated as paired end, with a non strand specific protocol. Samples denoted with '_3' were treated as paired end, with a strand specific protocol. 73
Figure 25 - Principal component analysis of gene expression levels in liver and testes tissues of 5 rodent species and 1 outgroup species 75
Figure 26 - Correlation plot of gene counts in liver and testes in rodent species using Spearman's Rank correlation 76
Figure 27 - Correlation plot of liver gene counts in inversions in the Muridae ancestor (left) and those not in inversions in the Muridae ancestor (right) using Spearman's Rank correlation 80
Figure 28 - Correlation plot of matched liver gene counts in inversions in the Muridae ancestor (left) and those not in inversions in the Muridae ancestor (right) using Spearman's Rank correlation 81
Figure 29 - Correlation plot of testes gene counts in inversions (left) and those not in inversions (right) using Spearman's Rank correlation 82
Figure 30 - Correlation plot of matched testes gene counts in inversions (left) and those not in inversions (right) using Spearman's Rank correlation 83
Figure 31 - GO terms enriched in the inversions between Muridae and Mus musculus with a p-value < 0.05 and FDR < 5% 84
Figure 32 - Mus musculus chromosomes with respect to reconstructed ancestors after manual merging of APCFs 120

List of Tables

Table 1 - Modes of speciation, adapted from [2] *14*

Table 2 - Sequence data for selected Rodentia species. *41*

Table 3 - Statistic of reconstructed ancestors using Figure 10 [A] as the evolutionary tree. *48*

Table 4 - Statistics of reconstructed ancestors using Figure 10 [B] as the evolutionary tree. *49*

Table 5 - Number and type of rearrangements between reconstructed rodent ancestors *64*

Table 6 - Rate of rearrangements by rearrangement type between reconstructed ancestors *65*

Table 7 - Orthologues genes missing from the Muridae ancestor reconstruction *78*

Table 8 - Gene counts for RNA-Seq data *121*

Abbreviations

ACS	Ancestral Contiguous Region
APCF	Ancestral Predicted Chromosome Fragment
Bp	Base pair(s)
CAR	Contiguous Ancestral Region
ChIP-seq	Chromatin immunoprecipitation sequencing
DNA	Deoxyribonucleic acid
EBR	Evolutionary breakpoint region
FDR	False discovery rate
FISH	Fluorescence *in situ* hybridization
Gbp	Giga base pair(s)
GO	Gene Ontology
GRAPPA	Genome Rearrangements Analysis under Parsimony and other Phylogenetic Algorithms
GRIMM	Genome Rearrangements In Man and Mouse
Kbp	Kilobase pair(s)
Mbp	Mega base pair(s)
MGR	Multiple Genome Rearrangements
MYA	Million years ago
NCBI	National Centre for Biotechnology Information
PCA	Principal Component Analysis
RNA	Ribonucleic acid
SD	Segmental duplication
SF	Syntenic fragment

SNP	Single nucleotide polymorphism
TAD	Topologically associating domains
TSP	Travelling Salesperson Problem
UCR	Ultra-conserved region
UCSC	University of California Santa Cruz

Abstract

One of the long-standing arguments in the area of evolutionary biology is the extent to which chromosomal rearrangements contribute to the process of speciation. The mammalian order Rodentia was used as a model to investigate the effects of chromosomal rearrangements on gene expression, using newly developed computational methods. Predicted ancestral karyotypes for 7 ancestors across 73 million years leading from the overarching Rodentia ancestor to mouse were reconstructed, which were then used to trace the number and type of rearrangements back through the lineage. Rodentia was found to be a highly rearranged order, with an average of 6.6 rearrangements per million years, higher than that seen in similar studies in birds and Eutherians. The ancestral reconstructions were also found to be highly fragmented, producing diploid numbers often double that in comparable cytogenetic predictions, suggesting that the reconstructions need further refinement to be representative. The effect of chromosomal rearrangements on gene expression was investigated using RNA-Seq data from liver and tissue, and the inversions identified from the ancestral reconstructions, due to their link to recombination suppression. Gene expression correlation was compared between species for gene orthologues found within inversions between mouse and the Muridae ancestor, compared to those not in inversions. A reduction of gene expression was seen in genes present in inversions, however this was found to be statistically insignificant. The results of this work do not indicate that speciation is driven by inversions in Rodentia, however it is believed that future work on the reconstructions, and greater understanding of the implication of the wider genome architecture on gene expression, may lead to a more complete picture. There are still many avenues for future work to investigate before chromosomal speciation can be ruled out in this instance.

Introduction

Speciation and Genome Evolution

The Earth is home to a vast array of life, in all sizes and shapes, found in every niche imaginable on the planet. The exact number of extant species is not exactly known, but is expected to be in the region of 8.7 million distinct species, of which only 1.2 million have been described by science [1]. Not to mention the millions of species which have risen and fallen over evolutionary time. Each of these species past and present have come to be due to the process of speciation, the process by which species arise. Some of the hypothesized speciation modes are listed in Table 1.

Table 1 - Modes of speciation, adapted from [2]

MODES OF SPECIATION

I. Classified by geographic origin of reproductive barriers
   - A. Allopatric speciation
     1. Vicariance
     2. Peripatric speciation
   - B. Parapatric speciation
   - C. Sympatric speciation

II. Classified by genetic and causal bases
   - A. Genetic divergence
     1. Genetic drift
     2. Peak shift
     3. Natural selection
   - B. Cytoplasmic incompatibility
   - C. Cytological divergence
     - *a) Polyploidy*
     - ***b) Chromosome rearrangement***
   - D. Recombinational speciation

The focus of this study is on the cytologic divergence mode of speciation, namely in relation to chromosome rearrangements. There are a number of models which have been described to explain the finer mechanics of chromosomal rearrangements potentially leading to speciation, which will be explained in greater detail later (see *Models of Chromosomal Speciation)*. Broadly speaking, chromosomal rearrangements are implicated in speciation due to heterozygotes containing one or more rearrangements leading to reproductive isolation either by reduced fertility and underdominance [3], or the reduction in gene flow due to suppressed meiotic recombination [4]. This would lead to subsets of populations accumulating chromosomal differences, which would eventually lead to two distinct populations of separate species incapable of producing viable hybrids.

Genomic rearrangements can also lead to changes, or interruption in the amount of gene expression [5]. Normal development of an individual not only relies on the presence of required genes, but also relies on these genes being expressed at the correct levels (gene dosage) at the correct times. Balanced chromosomal rearrangements such as reciprocal translocations and inversions do not modify the amount of genetic material, but they can change gene order. This could result in the deactivation of a gene, if the double stranded break were to occur within the gene itself. It could result in gene fusions where the double stranded breaks occur in two different genes, and then fuses them or their regulatory elements together [6]. It could also result in the disruption of gene regulation pathways, by interrupting the regulatory elements of the gene. Unbalanced chromosomal rearrangements such as deletions, unbalanced translocations, and duplications can cause aberrant gene expression due to incorrect gene dosage.

Types of Genomic Rearrangements

Changes in chromosome number can arise due to chromosome fusions and fissions. Chromosomal fusions involve the joining of two non-homologous chromosomes to form a new chromosome. One such example of a chromosome fusion is the telomeric fusion between hominoid ancestral chromosomes 2p and 2q in human to form chromosome 2 [7]. This has been evidenced by the discovery of telomeric sequences in band 2q13 [8] and a partly conserved ancestral centromere [9].

Chromosomal fission, inversely, involves the splitting of one chromosome into two separate non-homologous chromosomes. Human chromosomes 14 and 15 are the result of the fission of a hominoid ancestral chromosome 25 million years ago [10] mediated by segmental duplications (SDs) [11].

Chromosomal rearrangements occur due to double-stranded breakage in the DNA, followed by the joining of broken ends back together in a different way from the original gene order of the chromosomes. There are two main categories of rearrangement: balanced and imbalanced. Balanced rearrangements include inversions and translocations, these rearrangements do not result in a substantial change in the amount of DNA encoded by the chromosome (Robertsonian translocations do lose a small amount), but does change the gene order of the chromosome. Conversely imbalanced rearrangements, deletion and duplication, result in the gain or loss of genetic information. Where double-stranded breaks occur within genes, or within their supportive elements (promotors, enhancers etc) the break leads to a gene mutation.

Inversions occur when two double stranded breaks occur on the same chromosome, the portion of chromosome is then repaired, but in a different gene order, as shown in Figure 1. There are two main types of inversion, paracentric inversions which occur when the centromere is outside the inversion, and pericentric inversions where the centromere is within the inversion.

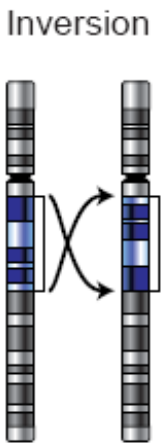


Figure 1 - Chromosomal inversion (image modified from 'Chromosomenmutationen' - Wikimedia commons)

Due to the balanced nature of inversion rearrangements, they tend to be viable, and do not often lead to phenotypic abnormalities, or have any clinical significance [12]. An example of this is the 12 Mb paracentric inversion in human chromosome 10 found in 0.2% of Swedish individuals, but with no consistent alteration in phenotype [13]. One disease which has been associated to inversions is Haemophilia A, where 42% of patients were found to have inversions within the Factor VIII gene [14] with inversions found both proximally and distally [15].

Inversion heterozygotes produce inversion loops during meiosis, and create deletion products during crossing-over, lowering the recombination frequency [16], and

reducing viability of zygotes [17]. The outcome is the same in both paracentric inversions and pericentric inversions, but the mechanisms that lead to it are slightly different, as summarised in Figure 2. In paracentric inversion heterozygotes a dicentric bridge is formed, with an acentric fragment. The acentric fragment is lost during anaphase, due to the lack of centromere, and the dicentric bridge is broken by tension, forming two deletion products. In pericentric inversion heterozygotes, crossing-over and separation occur as normal, however two of chromatids produced have a duplication in one region, and a deletion in another region, making the chromatid inviable due to genetic imbalance.

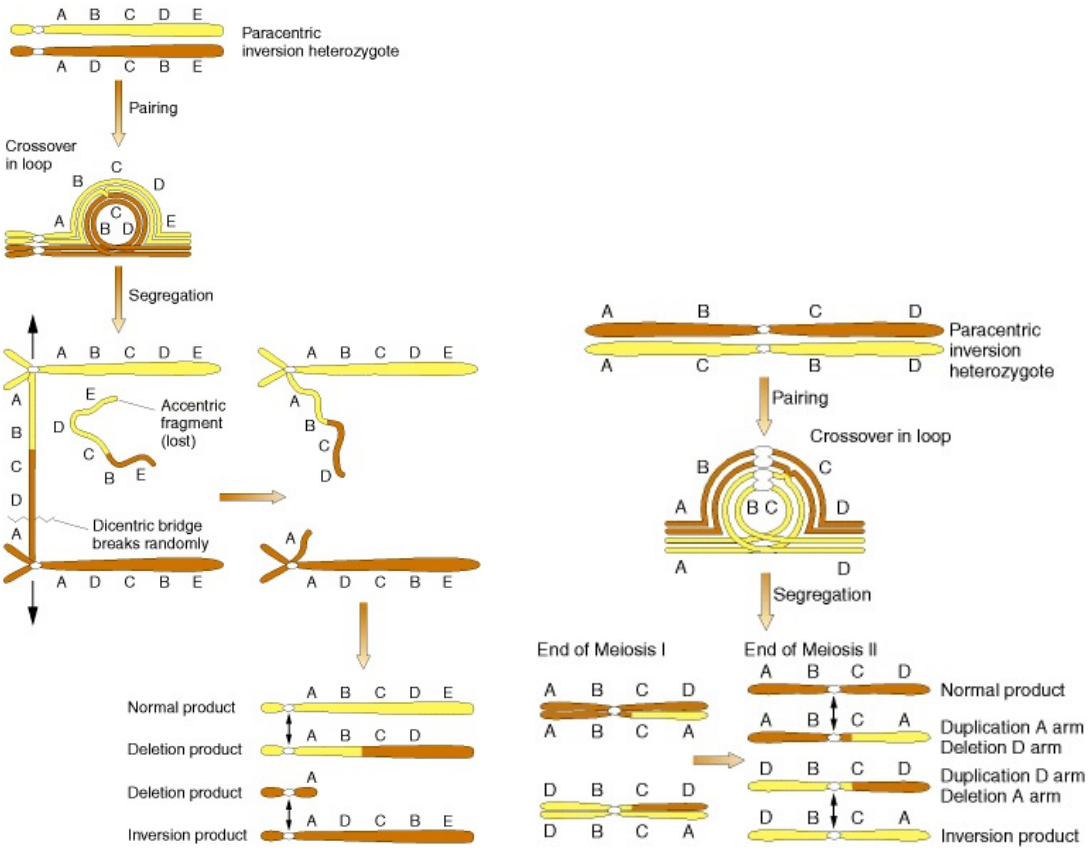


Figure 2 - Meiotic products resulting from a paracentric inversion heterozygote (left) and a paracentric inversion heterozygote (right) [17]

Translocations involve the transferring of sections of chromosome between non-homologous chromosomes, and occur in three different forms: reciprocal translocations (represented in Figure 3), non-reciprocal translocations and Robertsonian translocations. Both reciprocal and non-reciprocal translocations are balanced rearrangements, whereas Robertsonian translocations are imbalanced.

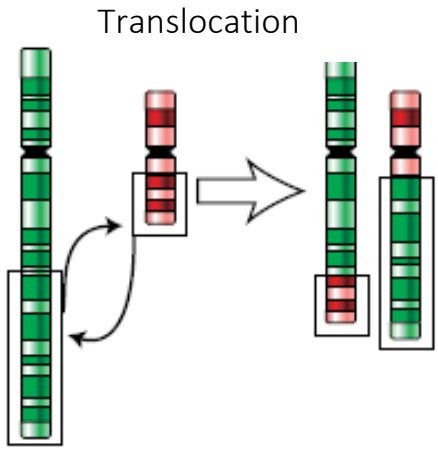


Figure 3 - Chromosomal translocation (image modified from 'Chromosomenmutationen' - Wikimedia commons)

Reciprocal translocations involve the mutual exchange of material between non-homologous chromosomes, whereas non-reciprocal translocations involve the transfer of a section of chromosome to another non-homologous chromosome without receiving anything in return. Robertsonian translocations only occur between acrocentric chromosomes. The long arm and short arm of the chromosome separate due to double-stranded breaks at the centromere, the long arms then fuse together to form one metacentric chromosome, the small arms and their small number of associated genes are lost.

Chromosomal deletions are where part of the chromosome is lost entirely, whereby two double-stranded breaks occur in the chromosome, the resultant broken segment is acentric so is lost during cell division, as it cannot be pulled towards a spindle pole during anaphase. The process results in a loss of a portion of the chromosome, as seen in Figure 4.

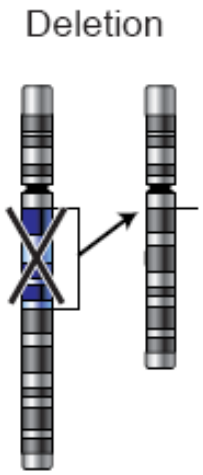


Figure 4 - Chromosomal deletion (image modified from 'Chromosomenmutationen' - Wikimedia commons)

The effect of the deletion is dependent upon the size. Smaller intragenic deletions which lead to the inactivation of just one gene can be viable, similar to variation caused by single nucleotide polymorphisms (SNPs).

Chromosomal duplications, sometimes referred to as insertions illustrated in Figure 6, are the opposite of deletions, in that they involve the gain of genetic information. Duplications where the duplicated region is adjacent to the original section are known as tandem duplications. Duplications where the duplicated region is located elsewhere on the chromosome, or in another chromosome entirely, are known as insertional duplications.

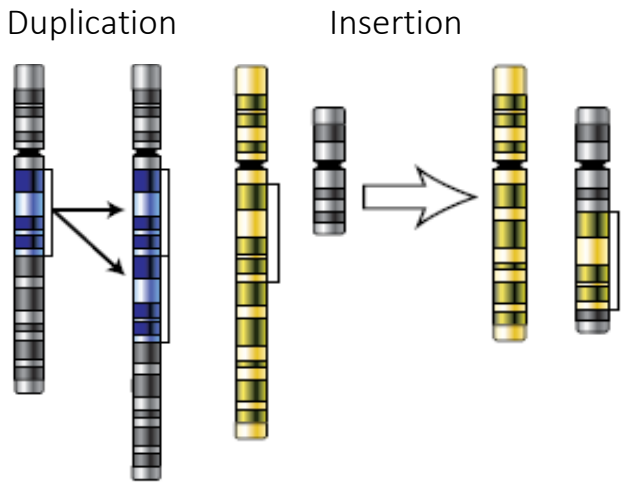


Figure 5 - Chromosomal duplication and insertion (image modified from 'Chromosomenmutationen' - Wikimedia commons)

Models for Chromosomal Speciation

There have been a number of models proposed over the years to explain chromosomal speciation, which fall into two broad categories for which there are multiple variants. These are the hybrid sterility models [3] and the suppressed recombination models [4] which are both illustrated in Figure 6. There have been a number of arguments for and against each of these models. Two recurring themes for controversy are the difficulties in fixing underdominant mutations in a population long enough to become genetically isolated, and the role of geographical isolation (or lack of) in tandem with chromosomal rearrangements.

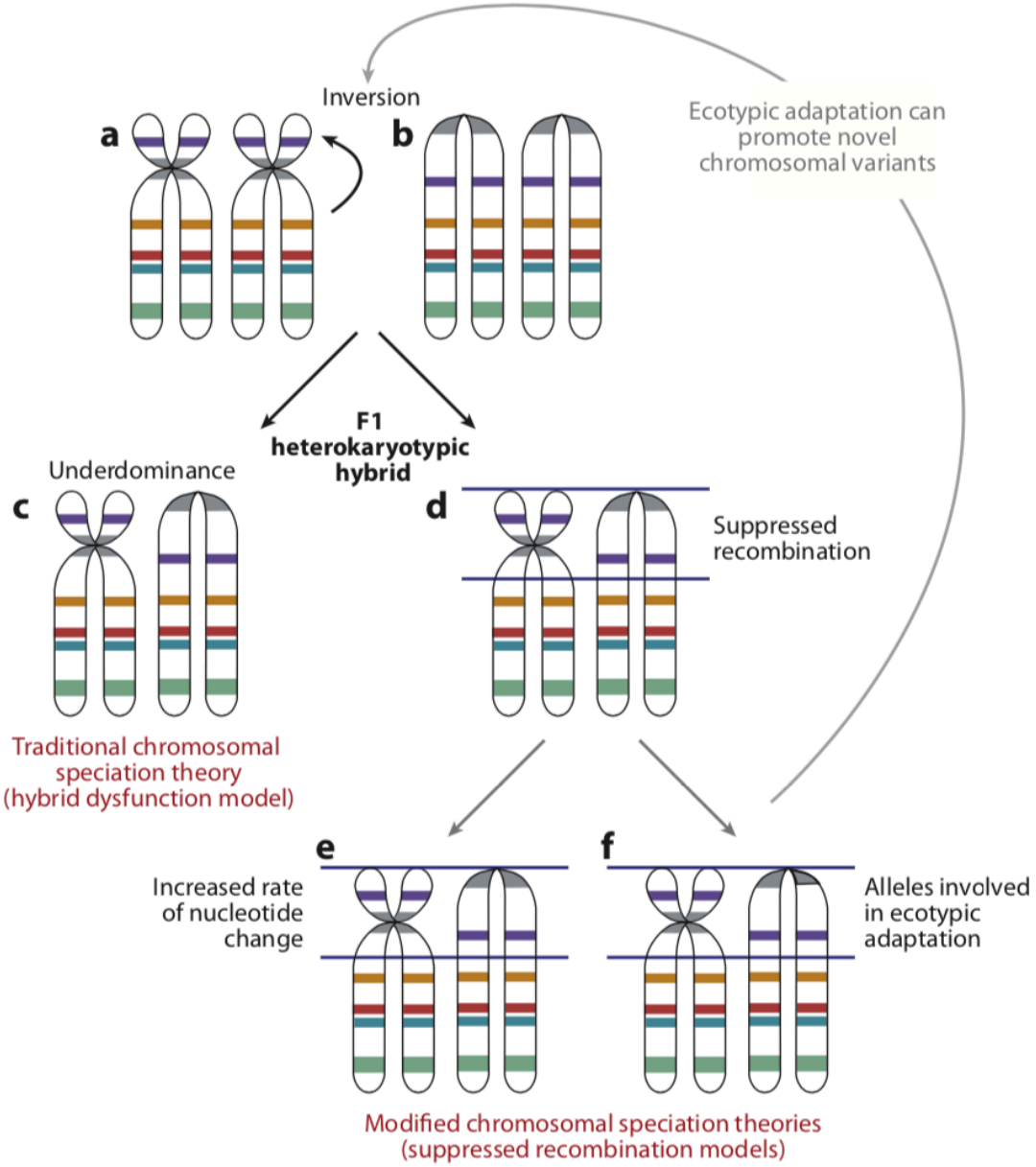


Figure 6 - Summary of the hybrid sterility model and recombination suppression model of chromosomal speciation [18]

Hybrid sterility models suggest that chromosomal rearrangements become fixed in a population, and that the recombination of these rearranged chromosomes in chromosomally heterozygous individuals reduces level of fitness or infertility (underdominance), introducing a barrier to geneflow [4,18]. There are a range of models which follow this structure with slight variants. The *Stasipatric model* suggests that a strongly underdominant chromosomal rearrangement becomes fixed in the

population by meiotic drive [3,4]. The *Chromosomal Transilience model* also suggests a strongly underdominant chromosomal rearrangement, but suggests that the fixation occurs due to inbreeding in an isolated population [4,19]. The *Chain* or *Cascade models* assumes an accumulation of weakly underdominant chromosomal rearrangements which combined give rise to reproductive isolation [3,4]. The *Saltational model* suggests that inbreeding within a founder population could lead to chromosomal breakage, the chromosomal rearrangements would then be fixed within that inbred population by genetic drift [4,20].

More recently proposed recombination suppression models suggest that chromosomal rearrangements reduce gene flow not by the reduction of fitness, but by suppressing recombination. It is thought that recombination suppression would result in either an increased rate of nucleotide change in the chromosome, or the preferential capture of alleles which confer a local adaptive advantage [21]. There have been studies to support recombination suppression, both directly [22,23], and indirectly through the genetic differentiation in the area surrounding inversion breakpoints [24].

Models for Genome Evolution

There have been a number of models proposed to try and explain the mechanisms that drive genomic rearrangements. The first model proposed was the *Random Breakage Model* [25,26] which was based on four assumptions:

1. Synteny of two or more markers in both species compared is presumptive evidence for linkage conservation.
2. Autosomal rearrangements fixed during evolution are distributed randomly throughout the genome.

3. Crossovers during recombination are distributed randomly throughout the genome.
4. Distribution throughout the genome of homologous markers is random and independent.

Studies involving genetic linkage maps found that spacing between markers was close to random [27], supporting the *Random Breakage Model.* Further to this, whole-genome sequence alignments were found to have synteny block length distributions [28] consistent with the *Random Breakage Model,* lending further support to the theory.

The later sequencing of the human [29] and mouse [30] genomes provided information that allowed this model to be called into question. Pairwise alignments between human and mouse found, using breakpoint graphs, that there were a much larger number of rearrangements found between the two species [31] than allowed for under the *Random Breakage Model*, in particular microrearrangements which had previously been ignored. In addition, breakpoints were mathematically found to be extensively reused in rearrangement “hotspots” [32]. The *Random Breakage Model* also does not account for ultra-conserved regions (UCRs) in the genome [33]. UCRs are regions of the genome which are highly conserved over a vast span of evolutionary time. UCRs clustered around vertebrate development have been conserved for 450 million years of vertebrate evolution, and often span hundreds of kilobases around target genes, some of them being >1000 bp in length [34].

These findings led to the proposal of the *Fragile Breakage Model* which postulates that the genome is a mosaic of fragile regions and solid regions, where breakpoint

regions occur largely within the short fragile regions of the chromosome, acting as rearrangement hotspots [32]. This model has been supported by a number of cytogenetic studies which have demonstrated the presence of evolutionary breakpoint regions (EBRs) within known fragile regions of the genome [35–38].

The *Intergenic Breakage Model* [39] argues that EBRs are not located in preferred sites, but actually occur randomly and that natural selection prevents unfavourable breakpoints that disrupt gene expression, so they develop in regions where there is not selection against them. They demonstrated this idea by artificially extending regulatory regions of genes, performed random microrearrangements and found that breakpoint reuse rate changes as the size of the regulatory region was increased [39]. Studies into EBR regions found that EBRs are underrepresented in genes [40] however are found to have a higher density in gene rich regions of the genome [40,41], with the hypothesis that EBRs correspond to areas of high transcriptional activity [40].

The *Integrative Breakage model* takes a multifactorial approach which accommodates observations made in the *Fragile Breakage model* and the *Intergenic Breakage model*. The model states that double stranded breaks occur in intergenic regions, that there are unstable genomic regions, and acknowledges the importance of chromatin conformation in the evolution of the genomic architecture [42]. The model also takes into account DNA sequence composition, the nucleome, and the effect on gene expression [42]. This theory is based on a range of observations surrounding the wider regulation of the genome, including: the presence of segmental duplications [43], tandem repeats [44], and transposable elements [45] at EBRs. It also observes what is currently known about the organisation of the genome within the nucleome, that

active euchromatin resides in the inner area of the nuclei, whereas the inert heterochromatin resides at the periphery of the nuclei [42,46]. Newly developed formation capture techniques will be able to lend more knowledge to the chromosomal rearrangement and gene expression debate by analysing the positions and frequency of positions of various loci within the cell.

Methods for Predicting Ancestral Karyotypes

In order to assess how chromosomal rearrangements may have led to speciation in a lineage, the evolutionary process of that lineage must first be inferred. Tracing the karyotype of an organism back to a common ancestor allows the evolutionary history of that lineage to be investigated, and comparisons between species to be made. It is an important area of phylogenomics and as such, there are a variety of methods that have been developed to infer ancestral karyotypes.

Cytogenetic Methods

Comparative cytogenetics was the first area of research to investigate the relationships of chromosomes between species and postulate as to their evolutionary history and ancestral karyotypes, with techniques such as zoo-FISH [47,48] and comparison of gene maps [49].

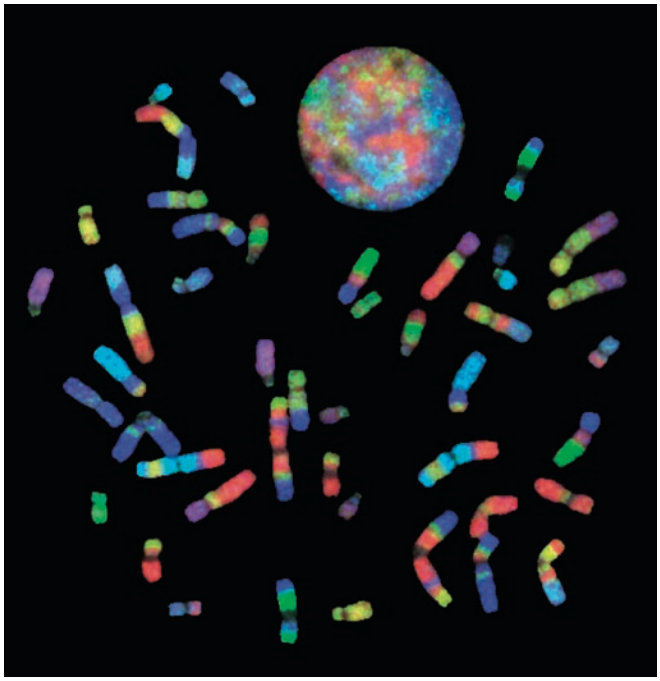


Figure 7 – Chromosome painting carried out on human chromosomes using chromosome-specific paint probes derived from gibbon chromosomes [50]

Cytogenetic methods have led to the ancestral chromosome predictions for a vast number of groups, including: the avian ancestor [51], the Xenartha ancestor [52], the Cetartiodactyla ancestor [48], the Eutherian ancestor [53], the Carnivora ancestor [54], the marsupial ancestor [55], and the ancestor to primates [56] amongst others. There are however limitations to cytogenetic methods. These limitations being a lack of resolution, which causes chromosome painting methods to miss intrachromosomal rearrangements, and a lack of evolutionary depth in reconstruction.

Computational Methods

With the increasing number of sequenced genomes, new computational methods have been developed to detect chromosome rearrangements and define ancestral karyotype configurations at a higher resolution. Computational approaches for predicting ancestral genomes follow two general approaches: the global parsimony method and the local parsimony method. The global parsimony method infers the

minimum number of chromosomal rearrangements to convert one genome order into another [57]. The local parsimony approach uses adjacencies between each branch of the phylogeny of modern genomes to predict the ancestral order and orientation with the most parsimonious outcome [58].

ANGES is a Python programme which tackles the problem with the local parsimony approach. It does this by detecting genome sections which have similar markers between each pair of species, which are used to derive weighted Ancestral Contiguous Sets (ACS). These ACS are then subset for those which satisfy a variety of the Consecutive-Ones Property [59] to produce Contiguous Ancestral Regions (CAR) [60]. InferCARs is another system which utilises the local parsimony approach, which takes nets from pairwise sequence alignments, and uses them to progressively construct orthology blocks, conserved segments, and finally CARs using adjacencies between species and concepts from graph theory. [58] Other methods which use variants of this method of using adjacencies between branches include ProCARs which progressively computes adjacencies, sub setting for non-conflicting ones and adding them in, without discarding false adjacencies in a single step [61] ,and GapAdj [62] which uses gapped adjacencies rather than direct adjacencies used in other methods [63] to create a more thoroughly constructed ancestral genome.

Tools which implement the global parsimony approach include ‘Genome Rearrangements Analysis under Parsimony and other Phylogenetic Algorithms’ (GRAPPA) and the Multiple Genome Rearrangement (MGR) algorithm. GRAPPA is a further development from BPAnalysis [64], the original tool which labels all internal nodes with gene orders, and then iterates through all potential outcomes using the

Travelling Salesman Problem (TSP) based on breakpoint distances. GRAPPA builds on this using algorithmic engineering [65] to speed up the process, as the computational complexity of BPAnalysis became exponentially greater with every added genome, and therefore impossible to use in many scenarios [66]. MGR considers inversions, translocations, fusions, and fissions based on genome rearrangement distance rather than breakpoint distances. It also allows for the analysis of both unichromosomal and multichromosomal genomes [67]. Multiple Genome Rearrangements and Ancestors (MGRA) [68] takes this a further step by utilising multiple breakpoint graphs compared to pairwise breakpoint graphs in MGR, making it faster and not requiring of the same amount of information in the input phylogenetic tree [68].

All of the previously mentioned computational approaches for ancestral reconstruction are hindered by a lack of chromosome level assemblies. DESCHRAMBLER [69] on the other hand, allows for the reconstruction of ancestral genomes using both chromosome and scaffold level assemblies, broadening the scope of potential research. At its inception DESCHRAMBLER was used to reconstruct 7 ancestral genomes from human to the Eutherian ancestor [69], 14 ancestors from zebra finch to the Avian ancestor [70], and 4 ancestors from cattle to the Cetartiodactyl ancestor [71].

Order Rodentia

Rodents, particularly the laboratory mouse, have a long established history of use within genetic and biomedical research [72], amassing a great amount of data on the species, and related species. Despite this, ancestral reconstructions of the Rodentia ancestor have not yet been produced computationally, nor their rearrangement history studied. Rodents are a rich source of study for evolutionary biology; for their diverse karyotypes, their phenotypic diversity, some rare adaptations that facilitate their success, and their sheer number in both population and species indicating their evolutionary success. Rodents are the largest group of mammals, represented by 2,285 currently recognised species [73], and both of the largest mammalian families, with Muridae and Cricetidae being found within the order [74]. They inhabit every continent on the planet apart from Antarctica, successfully surviving harsh environments with unique adaptations, such as the ability of the Arctic ground squirrel (*Spermophilus parryii*) to survive a core body temperature as low as -2.9°C during torpor [75]. Some species have very unique biology of great interest to many areas of research, such as the Naked mole-rat (*Heterocephalus glaber*) which is of interest due to its longevity [76–78], cancer resistance [79,80], anoxia resistance [81], and pain insensitivity [82].

The order Rodentia is divided into three major lineages which can be further sub-divided into seven major clades: the mouse-related lineage which contains Anomaluromorpha, Castoridae, Geomyoidea, and Myodonta [83], the squirrel-related lineage which contains Sciuriodea and Gliridae, and the guinea-pig related clade (Ctenohystricia) [84]. An evolutionary tree of the order is shown in *Figure 8*. Which one of these lineages represents the ‘root’ of the evolutionary tree, has been an area of

relative uncertainty. Studies on retroposon fixation found that eight orthologous retroposon elements and six indels were fixed in the common ancestor for the mouse-related lineage and the Ctenohystricia, suggesting that the squirrel-related lineage is the root of the Rodentia evolutionary tree [85]. There have been suggestions that Caviomorpha diverged before the separation of primates and artiodactyls [86,87], bringing into question the monophyly of the order. However, recent studies have strongly supported the case for monophyly [88–90].

Research into the karyotype of rodents has revealed a vast variety of diploid chromosome numbers, ranging from 2n=10 in an *Akodon* species [91] up to 2n=102 in *Tympanoctomys barrerae* (Red vizcacha rat), the largest known chromosome number in the kingdom Mammalia [92]. Two schools of thought have been put forward as to what mechanisms have led to this large genome size. First is that the red vizcacha rat is the first known mammal to demonstrate tetraploidy [93,94], possibly by whole genome duplication [95]. Later chromosome studies demonstrated only two copies of each chromosome [96], suggesting the second hypothesis that the large genome size could be explained by the amplification of repetitive sequences. A more recent study working on whole genome and whole transcriptome analyses, supports the repetitive sequences hypothesis, finding that 45.8% of the red vizcacha rat genome is made up of highly redundant sequences [97].

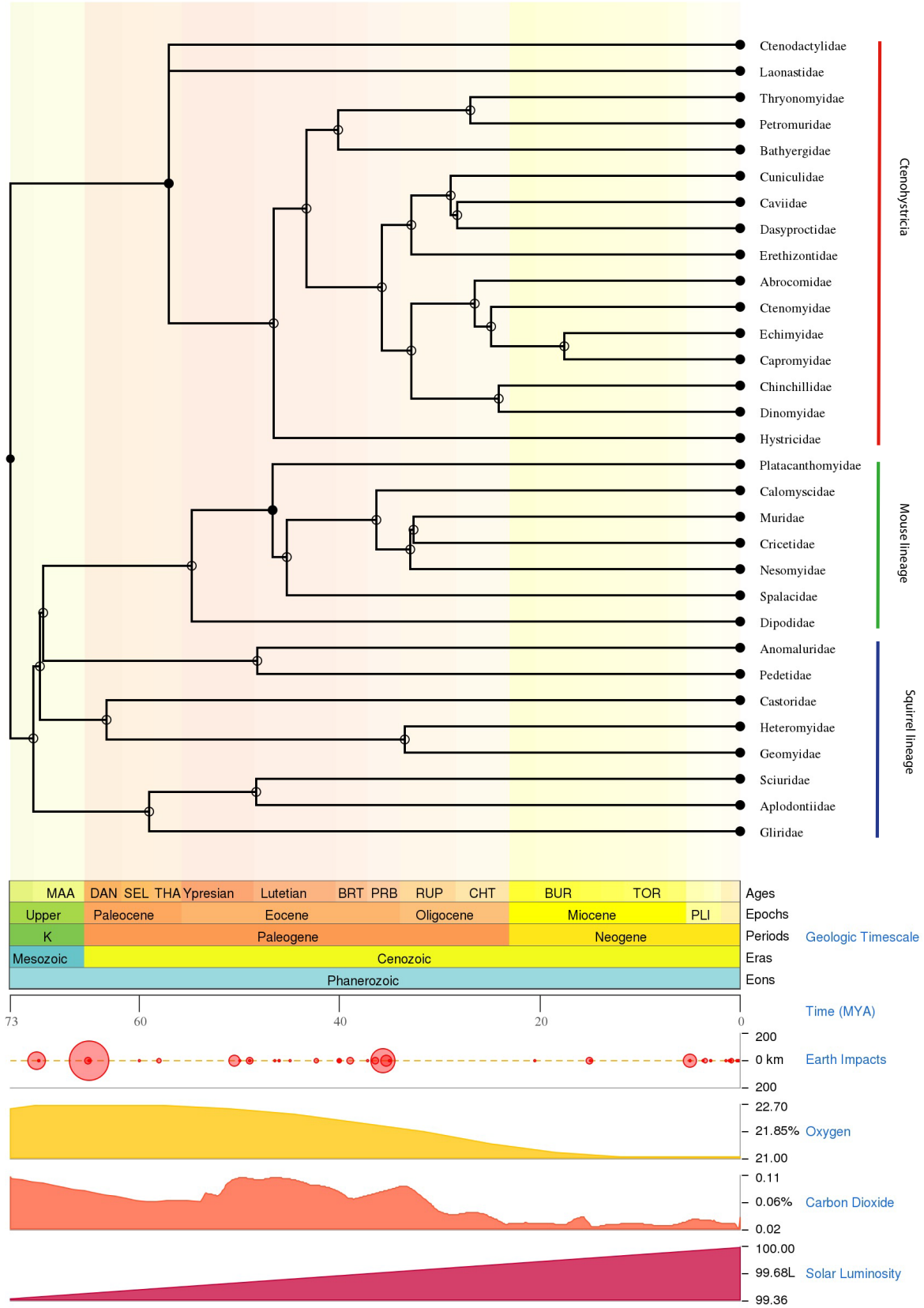


Figure 8 - Evolutionary tree of the order Rodentia, showing many of the major families [98]

Diploid number not only varies interspecifically, but is also found to vary intraspecifically; marked karyotypic variation has been observed in a number of rodent species. Diploid numbers of 2n = 16, 15 and 14 have been observed in *Akodon cursor* [99], with the 2n = 14 and 2n = 15 karyotypes for this species containing a large metacentric 1 chromosome. This metacentric chromosome arose due to pericentric inversions, and fusions of submetacentric chromosomes 1a and 1b, both of which can still be seen as distinct chromosomes in the 2n = 16 karyotype [100].

Project Aims

One of the outstanding questions in evolutionary biology, is the extent to which chromosomal rearrangements contribute to speciation [18]. To do this the evolutionary history of rearrangements in chromosomes must be deduced. This knowledge of rearrangements must then be combined with studies into gene expression to see how these rearrangements may be affecting gene expression in order to elucidate any possible implication in speciation. The aims of this project are therefore to:

1. Reconstruct ancestral karyotype predictions for ancestors leading from Rodentia ancestor to mouse.
2. Identify the number and type of rearrangements between each node of the evolutionary tree.
3. Obtain gene expression data for gene orthologues across study species.
4. Assess gene expression levels within rearrangements in different ancestral predictions.

Materials and Methods

Reconstruction of Rodentia Ancestors

Genomic Data

The genome assemblies of 14 Rodentia species and 3 mammalian outgroup species were downloaded from the National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov/assembly/?term=rodentia and DNA Zoo https://www.dnazoo.org/assemblies. Only chromosome assemblies or scaffold assemblies with a scaffold N50 value exceeding 3Mbp were included.

The species sourced from NCBI were House Mouse (*Mus musculus* – GRCm38.p6) [101,102], Norway Rat (*Rattus norvegicus* – Rnor_6.0) [103], Prairie vole (*Microtus ochrogaster* – MicOch1.0) [104], Chinese Hamster (*Cricetulus griseus* – CHOK1S_HZDv1) [105], Upper Galilee Mountains Blind Mole Rat (*Nannospalax galili* – S.galili_v1.0) [106], Lesser Egyptian Jerboa (*Jaculus jaculus* – JacJac1.0) [107], Ord's Kangroo Rat (*Dipodomys ordii* – Dord_2.0) [108], Thirteen-lined Ground Squirrel (*Ictidomys tridecemlineatus* – SpeTri2.0) [109], Yellow-bellied Marmot (*Marmota flaviventris – ASM367607v1)* [110], Domestic Guinea Pig (*Cavia porcellus* – Cavpor3.0) [111], Naked Mole-rat (*Heterocephalus glaber* – HetGla_female_1.0) [80], and Degu (*Octodon degus* – OctDeg1.0) [112]. The outgroup species sourced were Human (*Homo sapiens* – GRCh38.p12) [29,113], American pika (*Ochotona princeps* – OchPri3.0) [104], and Rabbit (*Oryctolagus cuniculus – OryCun2.0)* [114].

The species sourced from DNA Zoo were Canadian Beaver (*Castor canadensis)* [115–117]*,* Chinchilla (*Chinchilla lanigera*) [116–118] and the Damaraland mole-rat (*Fukomys damarensis)* [77,116,117]

Phylogenetic Tree Construction

Pairwise divergence times between *Mus musculus* and each of the study species were found using TimeTree [98]. The divergence times between species were then used to write a phylogenetic tree in Newick format and visualised using FigTree [119].

Pairwise Alignments

Five pairwise alignments between mouse and target species were downloaded from the University of California Santa Cruz (UCSC) genome browser https://genome.ucsc.edu/index.html : Mouse/human, mouse/guinea-pig, mouse/rabbit, mouse/rat, and mouse/squirrel [28,120,121]. The remaining genomes were prepared for alignment using Kent toolbox utilities [122,123] aligned to mouse using lastZ [124], using the parameters -minScore = 1000, -linearGap = medium, C = 0, E = 30, K = 3000, L = 3000, O = 400. The output of lastZ was then converted into chain and net files using Kent toolbox utilities [122,123].

The coverage of the nets of each species was calculated against the mouse genome as a target, to minimize the potential fragmentation introduced into the reconstruction of the ancestral karyotypes.

Reconstruction of Rodentia Ancestors

Ancestral predicted chromosome fragments (APCFs) were generated by the DESCHRAMBLER algorithm [69] using all species which had sufficient coverage against the reference species. The algorithm was executed using a syntenic fragment resolution of 300 kilobase pairs (Kbp) and a minimum adjacency score of 0.0001.

The output from DESCHRAMBLER produced a higher number of APCFs than the number of chromosomes suggested by previous studies [47,125–128]. Manual adjustments were made to the output of DESCHRAMBLER to merge together suitable APCFs in each of the reconstructed ancestors, using both the reference genome and the other reconstructed ancestors which were most closely related. This process was started on the Muridae ancestor using *Mus musculus* as a point of reference, before working back in evolutionary time using the closest related ancestors as a point of reference.

Identification of Chromosome Rearrangements

The Genome Rearrangements In Man and Mouse (GRIMM) [129] algorithm was used to determine the number and type of chromosomal rearrangements present at each stage leading from predicted Rodentia ancestor to *Mus musculus.*

Gene Expression Analysis

RNA-Seq Data

RNA sequencing (RNA-Seq) data was downloaded from the European Nucleotide Archive (ENA). Liver and testes data was downloaded for *Mus musculus* (PRJNA177791 [130]), *Rattus norvegicus* (PRJNA177791 [130]), *Fukomys damarensis* (PRJNA218853 [77]), *Cavia porcellus* (PRJNA385822 [131]), *Heterocephalus glaber* (PRJNA385839 & PRJNA385850 [131]), *Oryctolagus cuniculus* (PRJEB26840). Liver data was downloaded for *Nannospalax galili* (PRJEB17935 [132]).

RNA-Seq Alignment and Gene Counts

Downloaded RNA-Seq data was aligned using STAR aligner [133] to the respective genome assemblies (GRCm38.p6, Rnor_6.0, S.galili_v1.0, Cavpor3.0, DMR_v1.0, HetGla_female_1.0 and *OryCun2.0)* and genome annotations sourced from Ensembl. To minimise the effect of library preparation on the results, the alignment was carried out treating each sample as single end data.

Gene counts were quantified using the htseq-count function of the HTSeq framework [134] using a protocol which is not strand specific. The resultant counts were then normalised to correct for sequencing depth using the DESeq2 package [135] by utilising the ‘estimateSizeFactors’ and ‘counts’ functions which use the median of ratios method [136] of normalisation.

Filtering for Orthologues

Gene orthologues for *Mus musculus, Rattus norvegicus, Nannospalax galili, Cavia porcellus, Fukomys damarensis, Heterocephalus glaber* and *Oryctolagus cuniculus* were downloaded from Ensembl release 97 [137]. Resultant orthologues were filtered to use only one-to-one orthologues.

Correlation of Gene Expression

Mean expression for each gene in each species was calculated, in addition to mean expression for each gene across all species. Each set of tissue dataset was subset by species, and Spearman's rank correlation between species calculated in each tissue.

Gene Expression in Rearrangements

BiomaRt [137,138] was used to assign chromosome number, start position, and end position to each orthologue Gene ID. Start and end positions of each of the gene orthologues were then intersected with the start and end position of the syntenic fragments making up the reconstructed rodent ancestors, using the BEDTools intersect function [139].

Each tissue dataset was subset by species, and then further subset into two groups for genes within a given chromosomal rearrangement type, and those not present in the given chromosomal rearrangement type. Genes in rearrangements were matched to genes in non-rearrangements using MatchIT [140]. This allows us to ensure that correlations are not confounded by genes with extremes of mean gene expression values. For comparisons where the number of genes in one set was 10 times lower than in the other set, we matched genes one-to-one using 1000 permutations.

Expression of genes found in each type of rearrangement was compared to expression of genes in non-rearranged areas using a Wilcoxon rank sum test for paired data.

Gene Ontology (GO) Enrichment Analysis

A statistical overrepresentation test was carried out on PANTHER [141,142] using the genes found within rearrangements against the organism dataset for *Mus musculus* as the reference dataset. Terms with a p value of < 0.05 and a false discovery rate (FDR) of < 5% were considered to be significantly enriched.

Results and Discussion

Reconstruction of Rodentia Ancestors

Genome Selection and Alignment

There are 112 rodent sequence assemblies currently available on NCBI, ranging from contig assemblies to full chromosome assemblies. While the DESCHRAMBLER algorithm is able to use both chromosome and scaffold level assemblies, highly fragmented assemblies can introduce complications into the ancestral reconstruction [69], therefore only assemblies with an N50 scaffold size greater than 3 Mb were considered.

Phylogenetic diversity was also a consideration in the selection of assemblies; a range of assemblies from different Families were required to represent as much of the diversity in the Order as possible. To that end, assemblies that represented unique Families were included in the selection where genome quality met the aforementioned criteria. In instances where there were multiple assemblies for one family in particular, as was the case for the Muridae family, only the 1 or 2 highest quality assemblies were selected, to ensure that results were not skewed towards a particular grouping due to overrepresentation in the study.

The final selection of assemblies included 2 chromosome level rodent assemblies and 13 scaffold level rodent assemblies. Three assemblies were selected as outgroup species, the chromosome level Homo sapiens and Oryctolagus cuniculus assemblies, and the scaffold level Ochotona princeps assembly. Homo sapiens was selected as a high quality, distantly related outgroup, whereas Oryctolagus cuniculus and Ochotona

princeps were the only available assemblies for the Order Lagomorpha, the closest related Order to Rodentia. The sequence data for all the selected genome assemblies can be seen in Table 2

Table 2 - Sequence data for selected Rodentia species.

Organism	**Diploid Number (2n)**	**Family**	**Sequence Length (Gbp)**	**No of Scaffolds**	**N50 (Mbp)**
Mus musculus *	40 [143]	Muridae	2.8	336	52.6
Rattus norvegicus *	42 [143]	Muridae	2.9	1,395	15.0
Cricetulus griseus	22 [143]	Cricetidae	2.6	8,264	62.0
Microtus ochrogaster	54 [143]	Cricetidae	2.3	6,450	17.3
Nannospalax galili	-	Spalacidae	3.1	154,976	3.6
Jaculus jaculus	48 [144]	Dipodidae	2.9	10,898	22.1
Dipodomys ordii	72 [145]	Heteromyidae	2.2	65,193	11.9
Castor canadensis	40 [143]	Castoridae	2.5	6,496	136.7
Ictidomys tridecemlineatus	34 [146]	Sciuridae	2.5	12,483	8.1
Marmota flaviventris	42 [147]	Sciuridae	2.6	32,915	17.2
Fukomys damarensis	-	Bathyergidae	2.3	73,969	62.6
Cavia porcellus	64 [143]	Caviidae	2.7	3,144	27.9
Chinchilla lanigera	64 [143]	Chinchillidae	2.4	2,846	74.4
Heterocephalus glaber	-	Heterocephalidae	2.6	4,229	20.5
Octodon degus	58 [143]	Octodontidae	3.0	7,135	12.1
Oryctolagus cuniculus *	44 [143]	Leporidae	2.7	3,318	36.0
Ochotona princeps	68 [143]	Ochotonidae	2.2	10,421	26.9
Homo sapiens *	46 [143]	Hominidae	3.3	874	59.4

** denotes a species assembled to chromosome level, the remainder are assembled to scaffold level.*

Mus musculus was selected as the reference genome against which the other assemblies would be measured. This choice was based on the quality of the genome assembly, and placement within the Rodentia evolutionary tree. *Mus musculus* is assembled to chromosome level, and is both one of the most studied and highest quality mammalian genomes available due to being a model species for research across multiple disciplines. In addition to this, DESCHRAMBLER requires that the reference genome be a descendant of all the ancestors being reconstructed. With the assemblies selected, this allows for the reconstruction of 7 different ancestors, from the overarching ancestor for all of order Rodentia, right down to the Muridae ancestor of *Mus musculus* and *Rattus norvegicus.*

Pairwise alignments were either obtained from UCSC, or carried out using LastZ [124], and chain and net alignment files generated using Kent toolbox utilities [122,123] for each of the *Mus musculus* autosomes plus the X chromosome. The Y chromosome was omitted due to the difficulty in assembling it to a sufficient degree of quality, due to the enrichment of repeats and palindromes in the chromosome [148].

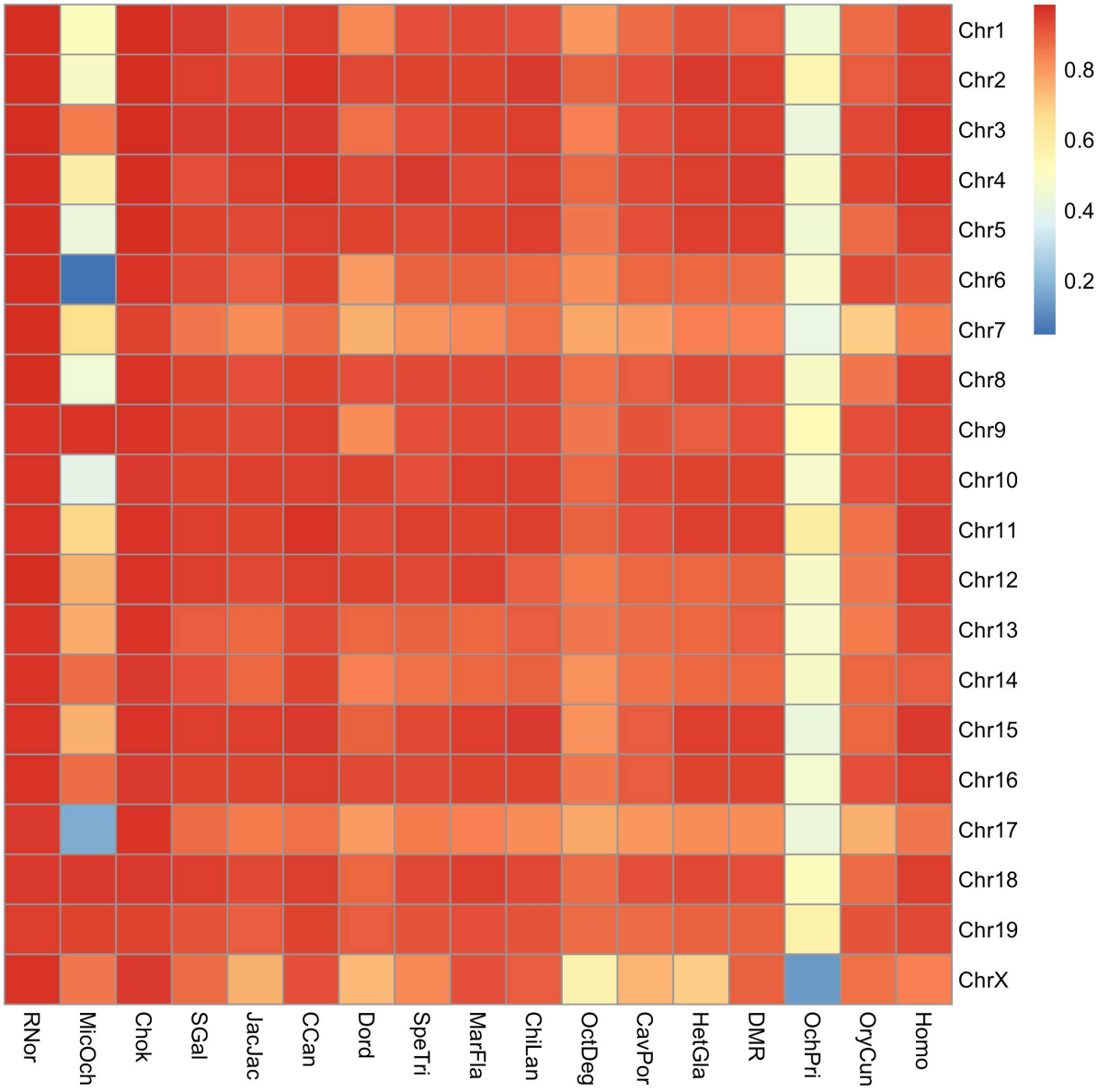


Figure 9 - Heatmap of the coverage of rodent net files of each chromosome as a percentage when compared to each chromosome of *Mus musculus*

Coverage of the net files for each chromosome produced by the pairwise alignments between the rodent species and *Mus musculus* were compared to the chromosomes of *Mus musculus,* the results of which can be observed in *Figure 9*, which shows the species used in pairwise alignments along the horizontal axis, the individual chromosomes along the vertical axis, and the coverage on a scale from 0 to 1 represented from poor coverage (blue squares) to strong coverage (red squares)

when compared to the same chromosome in *Mus musculus.* Having a high coverage of the reference genome ensures a more thorough coverage of the resultant predicted ancestral genomes, therefore only those species which had a greater than 80% coverage compared to *Mus musculus* were included in further analysis. *Microtus ochrogaster* was eliminated from further analysis at this stage, due to having only 63.07% coverage against *Mus musculus. Ochotona princeps* was also removed as an outgroup species at this stage due to having a net coverage of 46.26% against *Mus musculus*, despite being a chromosomal level assembly. The reduced coverage in both of these discarded species can clearly be seen in *Figure 9*, signified with yellow and blue boxes. Species that continued to the next stage of the study had a coverage ranging from 82.76% across all chromosomes in the Degu, to 97.37% across all chromosomes in the Norway Rat. The X chromosome had consistently the lowest coverage across all species.

Phylogenetic Trees

Pairwise divergence times were obtained from TimeTree [98] between *Mus musculus* and the study species, and used to produce a phylogenetic tree complete with divergence times as shown in *Figure 10* [A]. This phylogenetic tree suggests an earlier divergence of the guinea-pig related lineage (Ctenohystricia), with each of the species within this lineage having a pairwise divergence time of 73 million years ago (MYA). This is followed by the divergence of the squirrel related lineage (highlighted in red) 71 MYA, making it the nearest related lineage when compared to the mouse related lineage.

Due to the disparity between the tree proposed in *Figure 10* [A] and the body of work suggesting that the squirrel related lineage diverged prior to the Ctenohystricia [84,85,149,150], a second phylogenetic tree was produced swapping the positions of the squirrel related clade with that of Ctenohystricia, as shown in *Figure 10* [B].

The phylogenetic trees produced allows for the reconstruction of 7 different ancestors in the rodent lineage: Muridae, Eumuroidea, Muroidea, Myodonta, the ancestor for the mouse related lineage and the Rodentia ancestor in both scenarios. Using the tree shown in *Figure 10* [A] allowed for the reconstruction of the ancestor shared between the mouse related lineage and the squirrel related lineage. Using the tree shown in *Figure 10* [B] allowed for the reconstruction of the ancestor shared between the mouse related lineage and Ctenohystricia.

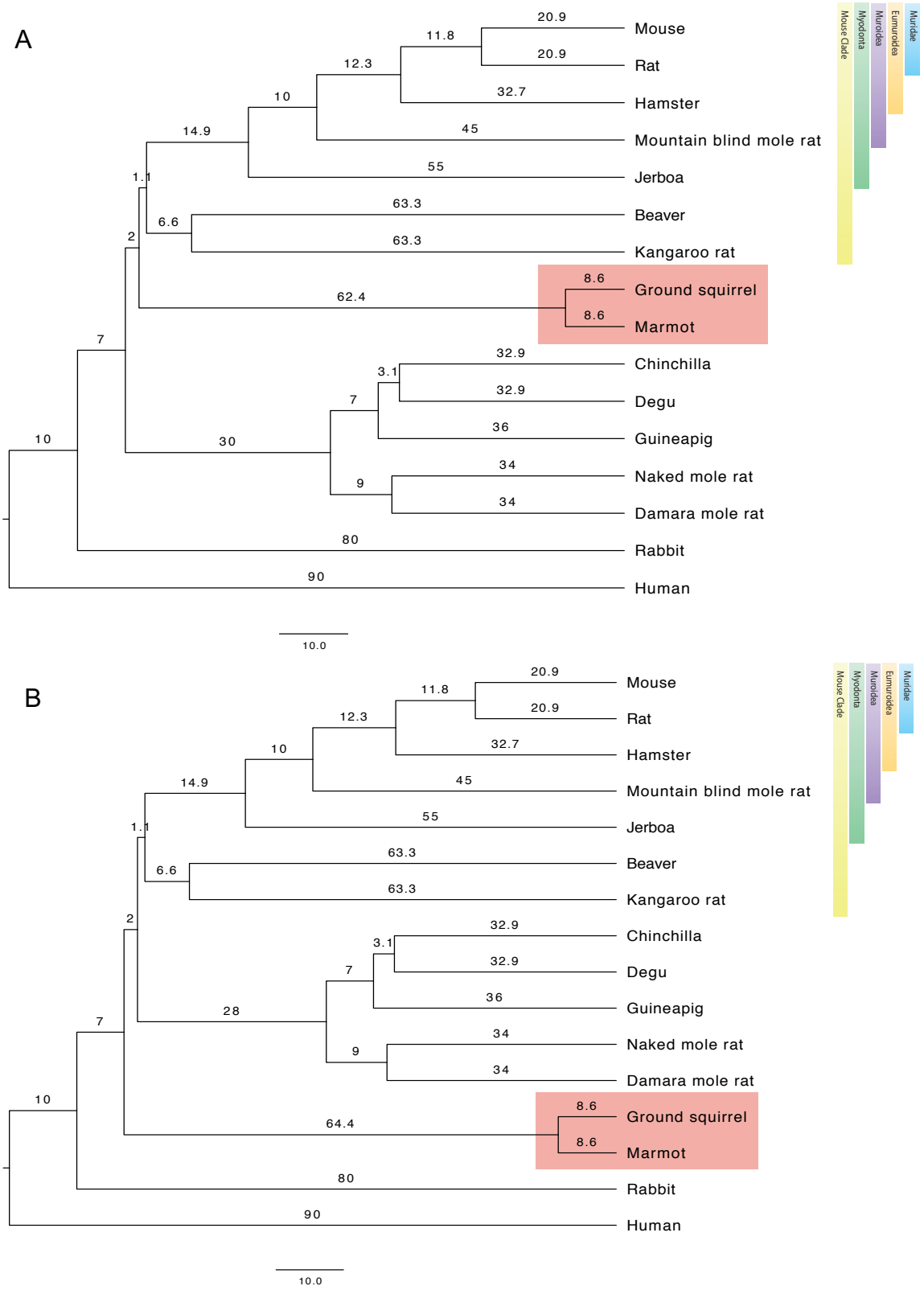


Figure 10 - Evolutionary trees for the Order Rodentia [A] where the squirrel related lineage (red) is the closest relation to the mouse related clade. [B] where Ctenohystricia is the closest relation to the mouse related clade, and the squirrel related lineage (red) is the 'root' of the rodent evolutionary tree.

Reconstructing Ancestral Predicted Chromosome Fragments

The DESCHRAMBLER algorithm [69] was used to reconstruct the APCFs for each of the 7 ancestors (Muridae, Eumuroidea, Muroidea, Myodonta, Mouse Clade, Mouse Clade + Squirrel Clade / Mouse Clade + Ctenohystricia, and Rodentia) for both constructed phylogenetic trees in *Figure 10*. The reconstructions generated from the phylogenetic tree with the squirrel lineage as the first lineage to diverge from the Rodentia ancestor produced APCF numbers ranging from 34 in Eumuroidea up to 78 in Myodonta *Table 3*, with coverage of the *Mus musculus* genome ranging from a low of 84.64% in the Rodentia ancestor up to 92.78% in the Muridae ancestor *Table 3*.

Of the 62 APCFs generated for the Muridae genome, 12 of these consisted of syntenic fragments which were shared between mouse and rat but not any of the ancestors. Higher numbers of APCFs were also observed in the Muroidea and Myodonta ancestors. This could be a reflection of the inclusion of *Nannospalax galili* in these reconstructions as an ingroup. *Nannospalax galili* was the most poorly assembled genome used in this study, with a scaffold number of 154,976, more than double that of the next most fragmented assembly used in the study. Having this poorly constructed assembly as an ingroup without the balance of the less fragmented assemblies representing the squirrel-related lineage and Ctenohystricia, which are outgroups at this stage in the reconstruction, may have made it harder for the algorithm to place synteny fragments efficiently.

Table 3 - Statistic of reconstructed ancestors using Figure 10 [A] as the evolutionary tree.

Ancestor Name	Total length of APCFs (bp)	Coverage of mouse genome (%)	No. APCFs	No. SFs
Muridae	2,443,617,095	92.78	62	483
Eumuroidea	2,406,220,951	91.36	34	567
Muroidea	2,420,519,854	91.90	60	1,523
Myodonta	2,383,577,826	90.50	78	1,673
Mouse Clade	2,347,559,341	89.13	49	1,853
Mouse Clade + Squirrel Clade	2,313,668,409	87.85	59	2,058
Rodentia	2,229,265,833	84.64	53	2,270

The reconstructions generated from the phylogenetic tree with Ctenohystricia as the first lineage to diverge from the Rodentia ancestor produced APCF numbers ranging from 34 in Eumuroidea up to 76 in Myodonta *Table 4*. Myodonta had 2 fewer APCFs, the mouse clade had 1 fewer APCF and Rodentia had 1 more APCF when compared to the results obtained from the previous phylogenetic tree. The greatest difference found between the two scenarios was the number of APCFs found in the ancestor of the Mouse clade and Ctenohystricia, which was 10 fewer than found for the ancestor of the mouse clade and squirrel clade in the previous phylogenetic tree. Coverage compared to the genome of *Mus musculus* ranged from a low of 84.99% in Rodentia to a high of 92.78% in Muridae *Table 4*, despite the coverage of Rodentia being higher when compared to the previous phylogenetic tree, the coverage of the Myodonta and Mouse clade ancestors were lower.

Table 4 - Statistics of reconstructed ancestors using Figure 10 [B] as the evolutionary tree.

Ancestor Name	Total length of APCFs (bp)	Coverage of mouse genome (%)	No. APCFs	Merged APCFs	No. SFs
Muridae	2,443,617,095	92.78	62	48	483
Eumuroidea	2,406,220,951	91.36	34	34	567
Muroidea	2,420,519,854	91.90	60	55	1,523
Myodonta	2,379,430,745	90.34	76	69	1,669
Mouse Clade	2,343,412,260	88.98	48	45	1,849
Mouse Clade + *Ctenohystricia*	2,276,768,990	86.45	49	48	2,131
Rodentia	2,238,577,751	84.99	54	52	2,283

The reduced fragmentation of resultant APCFs using the phylogenetic tree in *Figure 10* [B] shows that the DESCHRAMBLER algorithm was more able to resolve the genomic data with the topology of the phylogenetic tree, one of the elements that DESCHRAMBLER relies on [69]. This supports the argument for the squirrel related lineage being the first group to diverge from the Rodentia ancestor, therefore the phylogenetic tree represented in *Figure 10* [B] and the results from DESCHRAMBLER summarised in *Table 4* will be used to identify the chromosomal rearrangements.

The output from DESCHRAMBLER in both scenarios produced a higher number of APCFs than the number of chromosomes suggested by previous studies [47,125–128], which is most likely attributable to the fragmented nature of the predominantly scaffold level assemblies used throughout this study. Manual adjustments were made to the output of DESCHRAMBLER to merge together suitable APCFs in each of the reconstructed ancestors, using both the reference genome and the other reconstructed ancestors which were most closely related. This process was started on

the Muridae ancestor using *Mus musculus* as a point of reference, before working back in evolutionary time using the closest related ancestors as a point of reference.

After the manual adjustments made to the DESCHRAMBLER output the Muridae ancestor was reduced by 14 APCFs to a total of 48. The Muroidea ancestor was reduced by 5 APCFs to a total of 55, Myodonta reduced by 7 APCFs to a total of 69, Mouse clade reduced by 3 APCFs to a total of 45 APCFs, Mouse clade + Ctenohystricia reduced by 1 APCF to a total of 48, and Rodentia reduced by 2 APCFs to a total of 52 APCFs.

Myodonta and Muroidea remain having the highest number of APCFs, as before this could be a reflection of the highly fragmented *Nannospalax galili* being included as an ingroup in these reconstructions, with the better constructed squirrel related and Ctenohystricia assemblies still being outgroups at this stage.

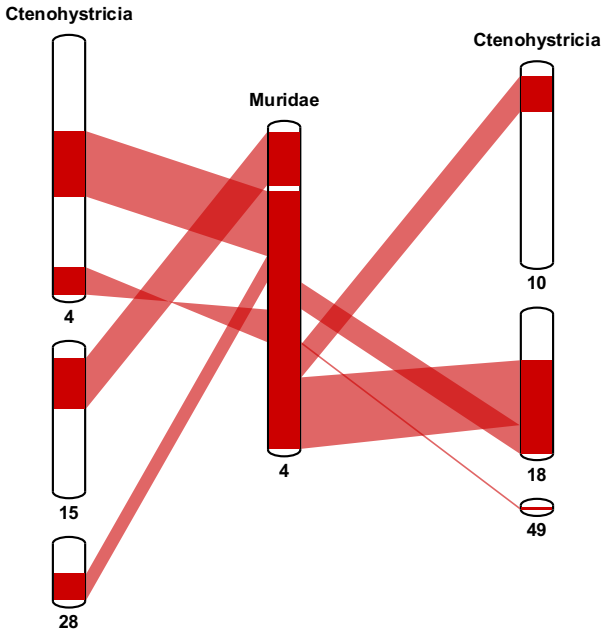


Figure 11 - mySynteny view of syntenic blocks shared between Muridae ancestor APCF4 and various APCFs in the Ctenohystricia ancestor

The reconstructed ancestors were visualised in the Evolution Highway format (see Appendix 1) and in mySynteny Portal [151], an example showing the syntenic relationship between Muridae APCF4 and the Ctenohystricia ancestor can be seen in Figure 11. The relationships between ancestors and ancestors, and ancestors and mouse are viewable at elii.net/rodentSynteny. Variability between chromosomes is seen in the level of fragmentation and the number of rearrangements leading to each of them. Mouse chromosome 15 *Figure 12* is an example of one of the less fragmented chromosomes, comprising of a maximum of 4 synteny blocks in any given ancestor. It is generated by a number of inversions, and a couple of fusions from Muroidea to Eumuroidea.



Figure 12 - *Mus musculus* chromosome 15 in relation to the APCFs of each of the predicted ancestors in the Evolution Highway format. Blue and pink blocks represent syntenic fragments in "+" (blue) and "-" (pink) orientation. The number in the block represents the APCF reference number.

Mouse chromosome 17 *Figure 13* on the other hand is an example of one of the more fragmented chromosomes, comprising of a greater number, but smaller in size, synteny blocks, with missing information interspersed between them. Between the 7 ancestors, Muridae and Eumuroidea share large sections of homology. Mouse clade, Mouse clade + Ctenohystricia and Rodentia also share large sections of homology. Muroidea and Myodonta share some sections of homology, though less so than the two previous groupings.

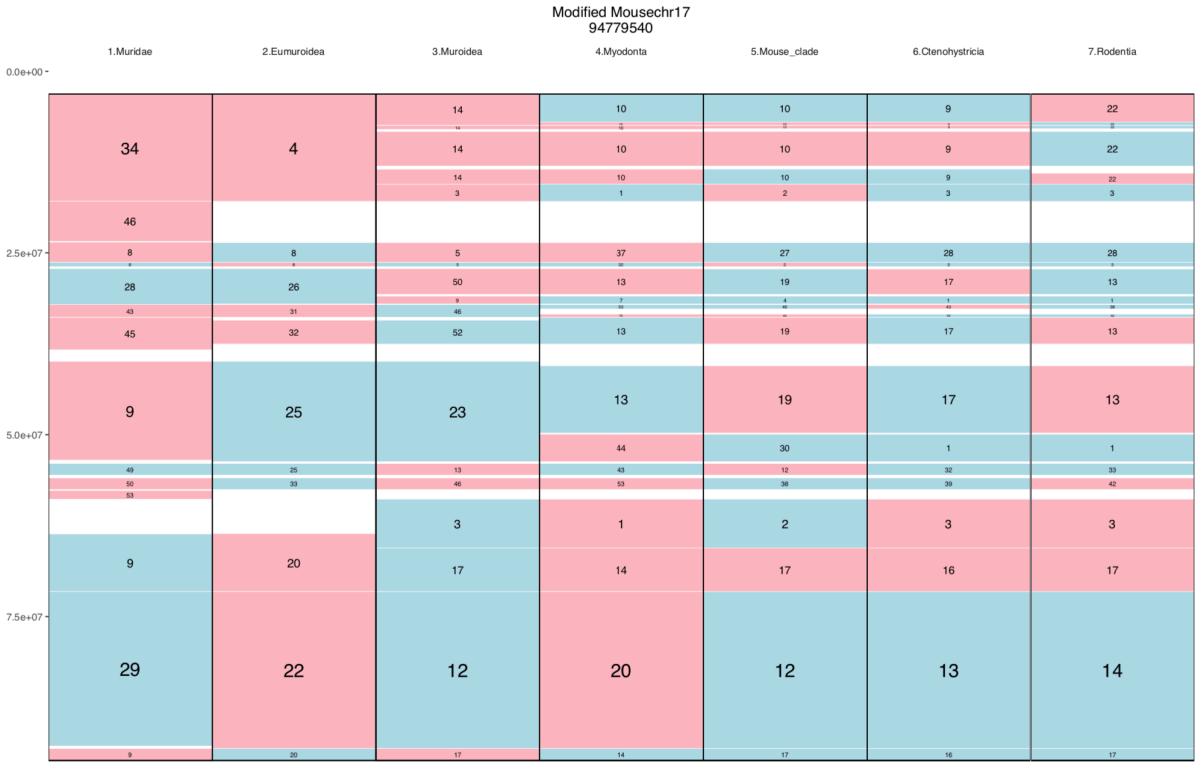


Figure 13 - *Mus musculus* chromosome 17 in relation to the APCFs of each of the predicted ancestors in the Evolution Highway format. Blue and pink blocks represent syntenic fragments in "+" (blue) and "-" (pink) orientation. The number in the block represents the APCF reference number.

Comparison to Cytogenetic Studies – Muridae

Chromosome painting studies have produced a predicted ancestral karyotype for the Muridae ancestor of 2n = 50 [126], summarised in Figure 14 which shows the karyotype in segments homologous to the mouse genome. This estimation of the karyotype ranges up to a high of 2n = 56 under certain interpretations of the homologous regions in mouse chromosomes 4, 5, and 10 [126].

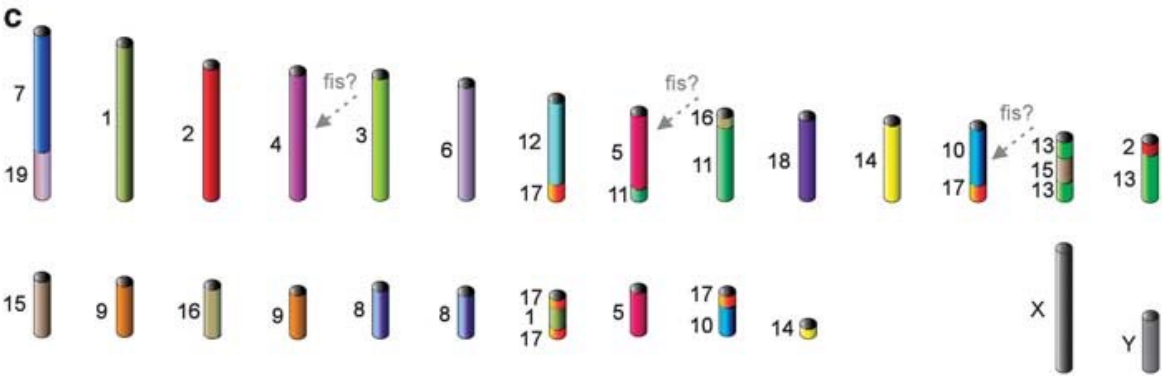


Figure 14 - Ancestral karyotype for the ancestor of Muridae. Different colours correspond to separate mouse chromosomes [126]

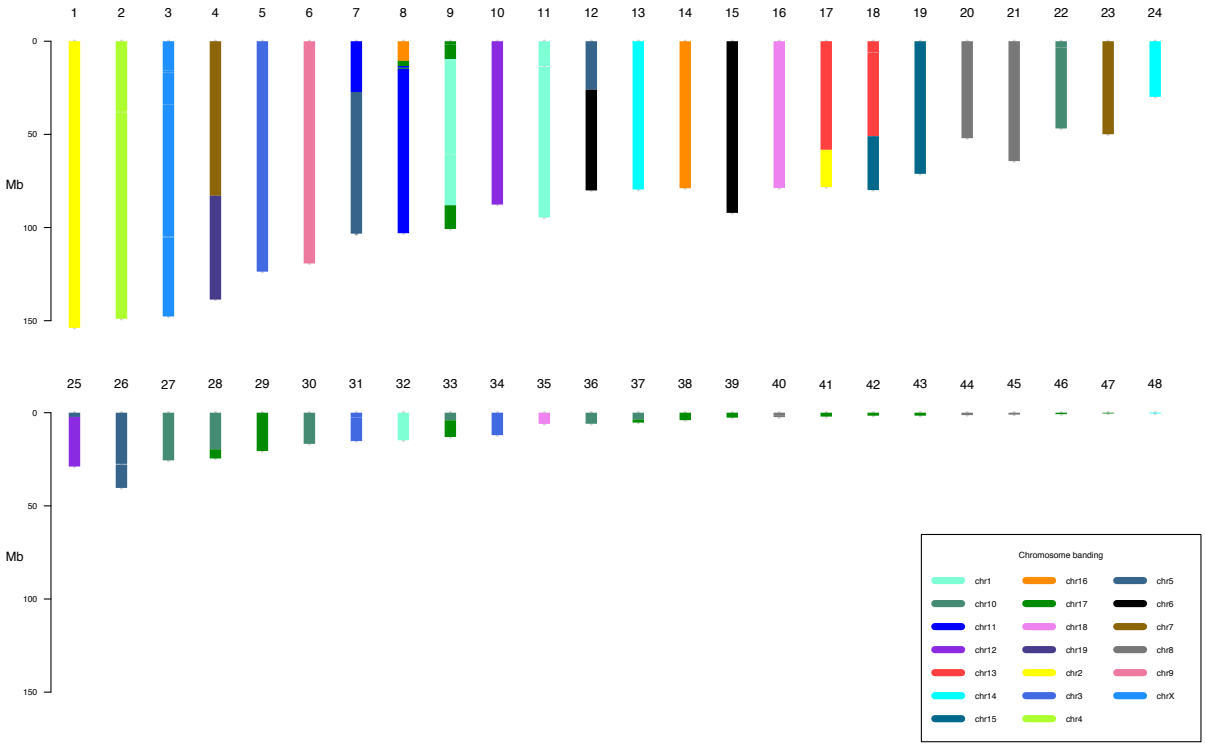


Figure 15 - Ideogram of APCFs for Muridae ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from mouse chromosomes

The APCFs of the Muridae ancestral prediction produced by DESCHRAMBLER are summarised in Figure 15 as ideograms made up of the homologous regions with the mouse genome. The number of APCFs predicted by this method is much higher than the 2n = 50 predicted by chromosome painting, at 2n = 96. Even for a family that has such a range of diploid numbers – from 2n = 14 in *Taterillus tranieri* to 2n = 74 in *Gerbillus latastei* [126] – this is an unrealistically high diploid number. This is a result of the fragmentation of the results; of the 48 APCFs generated, 19 of them were smaller than the smallest mouse chromosome.

Despite the fragmented nature of the computational reconstruction, of those larger reconstructed APCFs, 16 of them share similarities with the results from the

chromosome painting. The associations which are shared between the two predictions are MMU 2, 4, X, 3, 11/5, 1, 14, 16, 6, 18, 13/2, 15, 8, 8, 14, and 10/17. The association MMU 17/1/17 is found in both reconstructions, however the size of the chromosome 1 section appears to be larger in the computational reconstruction, however as both are ideograms, and the chromosome painting prediction does not offer sizes, it is hard to substantiate this, or assess significance.

There are also a number of ways in which these two predicted karyotypes differ. The association MMU 7/19 seen in Figure 14 is found in APCF4, however in the computational reconstruction there is also a separate region of MMU 7 seen in APCF23. There are two predicted chromosomes in Figure 14 which have the association MMU 9, however in the computational reconstruction MMU 9 is together in one APCF6. APCF8 has an association of MMU 16/17/11/17/11 which is not found at all in the chromosome painting based reconstruction, chromosome 17 has been consistently the most fragmented chromosome to work with, so this could be as a result of DESCHRAMBLER trying to place this highly fragmented chromosome. It could also be the case that the chromosome painting lacks the resolution to detect the very small sections of chromosome 17, and this prediction is actually related to the MMU 16/11 association seen in Figure 14. The association MMU 12/17 is not in the computational reconstruction, however APCF10 consists of a section of chromosome 12, and there are a number of small fragments of chromosome 17, which could potentially be associated. Finally, there is an association of MMU 13/15/13 in Figure 14, however only MMU 13/15 in the computational reconstruction, and no remaining APCFs consisting of chromosome 13 segments.

Comparison to Cytogenetic Studies – Eumuroidea

Chromosome painting studies between mouse (*Mus* musculus), golden hamster (*Mesocricetus* auratus), and the Chinese hamster (*Cricetulus* griseus) have produced a predicted ancestral karyotype for the Muridae ancestor of 2n = 48 [128], summarised in Figure 16 which shows the karyotype in segments homologous to the mouse and golden hamster genomes.

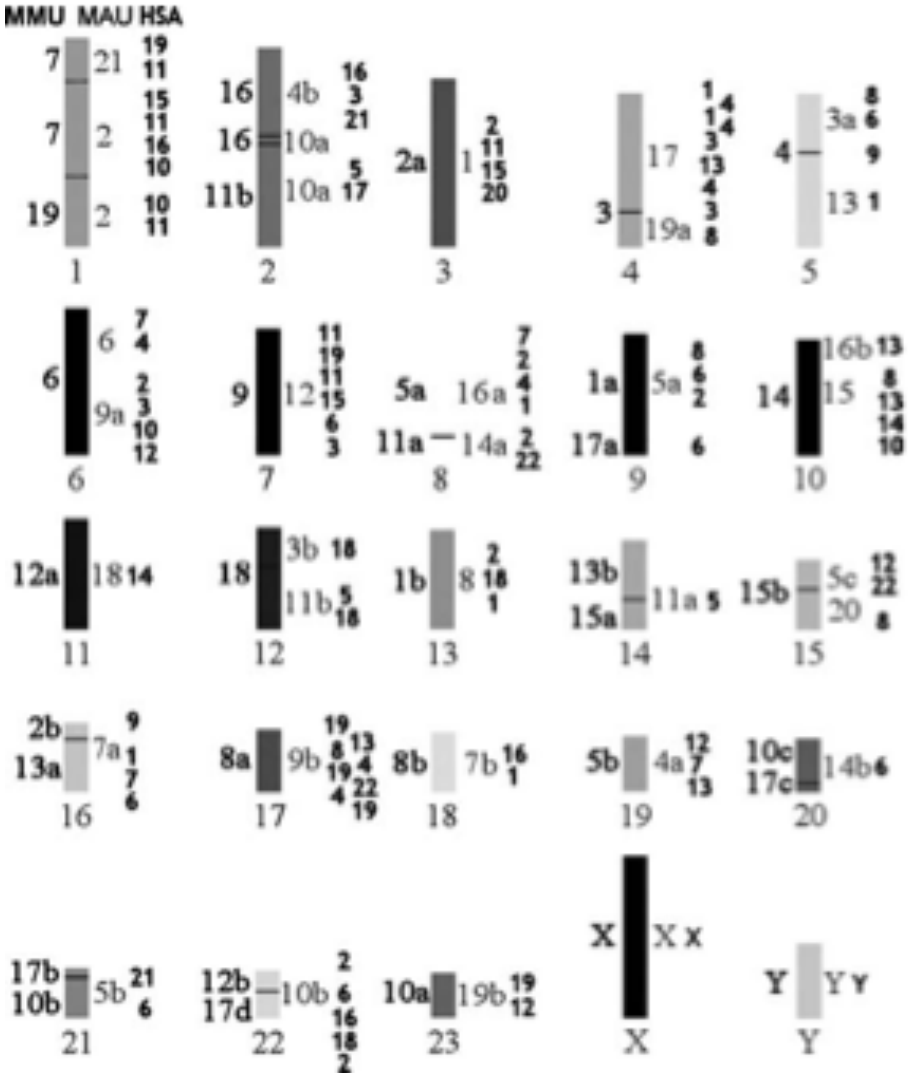


Figure 16 - Ancestral Eumuroidea karyotype. The homologies of mouse (MMU) chromosomes are shown to the left of the ideogram, and the homologies of golden hamster (MAU) chromosomes are shown to the right of the ideogram. Number in bold represent possible homology with human chromosome fragments [128]

The APCFs of the Muridae ancestral prediction produced by DESCHRAMBLER are summarised in Figure 17 as ideograms made up of the homologous regions with the mouse genome. The number of APCFs predicted by this method is higher than the 2n = 48 predicted by chromosome painting, at 2n = 68. Despite this being the predicted ancestral karyotype that was the least fragmented, there were still 8 APCFs constructed which were smaller in size than the smallest mouse chromosome.

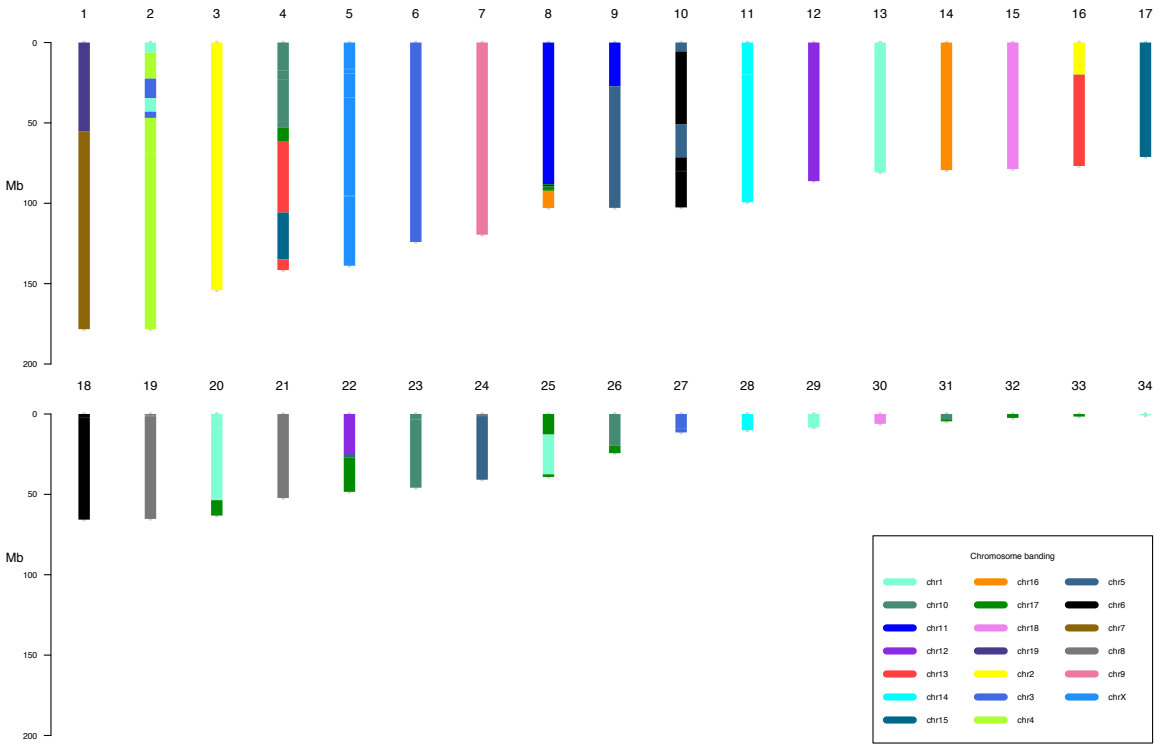


Figure 17 - Ideogram of APCFs for Eumuroidea ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from mouse chromosomes

Despite the fragmented nature of the computational reconstruction, of those larger reconstructed APCFs, 16 of them share similarities with the results from the chromosome painting. The associations which are shared between the two predictions are MMU 19/7, 2, X, 3, 9, 11/5, 14, 12, 1, 18, 8, 1/17, 8, 10, 10/17, and 17/10. There

were also a number of differences between the two predictions. Figure 16 shows a predicted chromosome made up of MMU 6, whereas in the computational prediction Figure 17 chromosome 16 is split between APCF18, and APCF10, which is also associated with MMU5. Figure 16 also shows an association of MMU 12/17, in Figure 17 this is MMU 12/5/17. The section of chromosome 5 within this predicted chromosome is likely too small to have been picked up by chromosome painting, which has a lower resolution than computational methods like DESCHRAMBLER.

Comparison to Cytogenetic Studies – Muroidea

Chromosome painting studies have produced a predicted ancestral karyotype for the Muroidea ancestor of 2n = 52 [126], summarised in Figure 18 which shows the karyotype in segments homologous to the mouse genome.

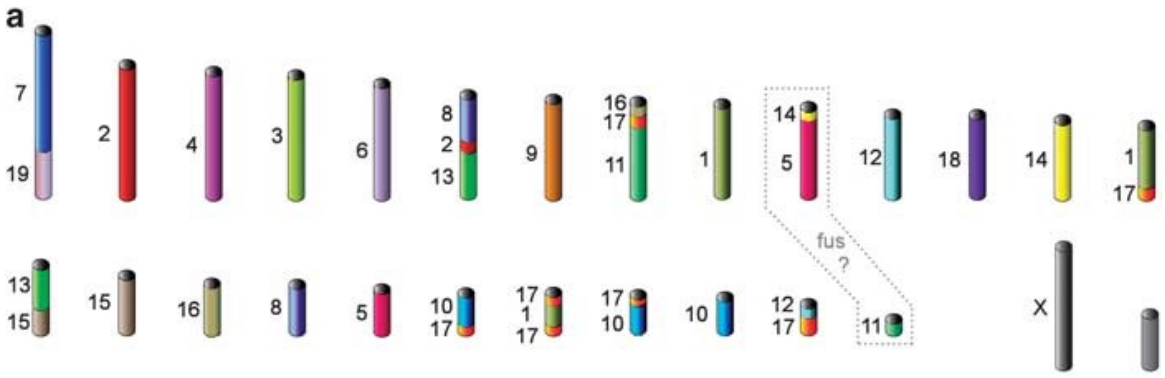


Figure 18 - Ancestral karyotype for the ancestor of Muroidea. Different colours correspond to separate mouse chromosomes [126]

A previous chromosome painting study based on 20 different rodent species from the families *Allocricetulus, Calomyscus, Cricetus, Cricetulus, Mesocricetus, Peromyscus, Phodopus* and *Tscherskia* produced a predicted ancestral karyotype for the Muroidea ancestor of 2n = 48 [125]. This is summarised in Figure 19 which shows the karyotype of the predicted Muroidea ancestor in segments marked with the homology to mouse

and golden hamster. The prediction of 2n = 52 is likely to be a more accurate prediction due to there being a greater representation of different families, rather than an overrepresentation of hamster species [126].

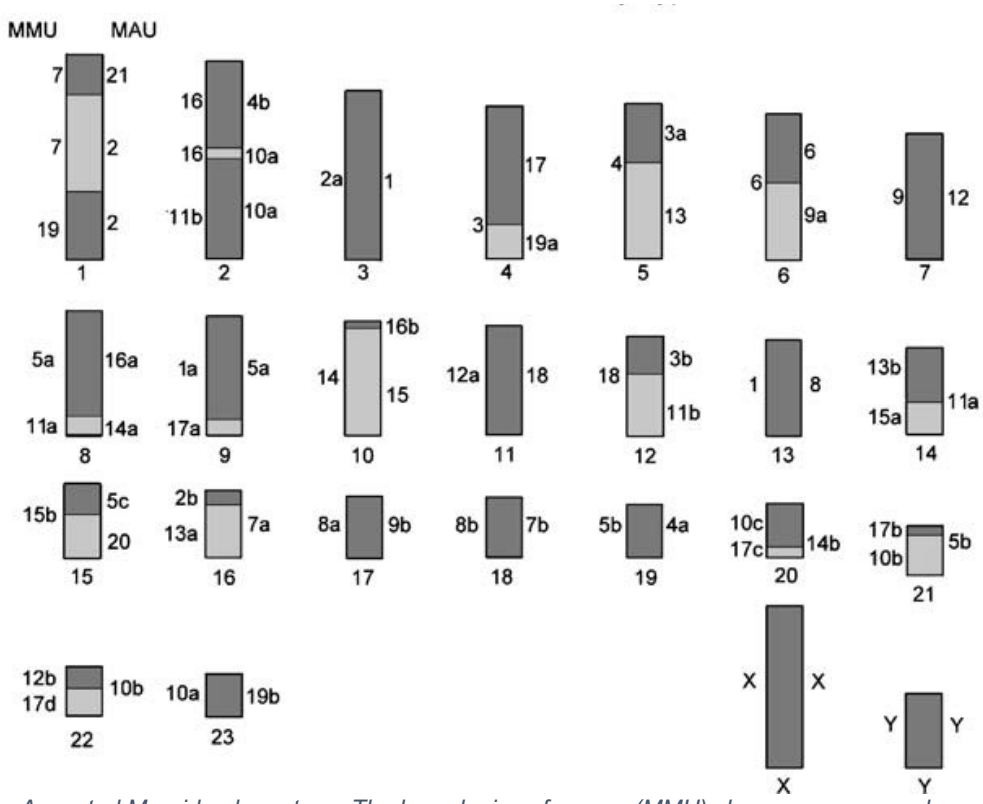


Figure 19 - Ancestral Muroidea karyotype. The homologies of mouse (MMU) chromosomes are shown to the left of the ideogram, and the homologies of golden hamster (MAU) chromosomes are shown to the right of the ideogram.

The APCFs of the Muroidea ancestral prediction produced by DESCHRAMBLER are summarised in Figure 20 as ideograms made up of the homologous regions with the mouse genome. The number of APCFs predicted by this method is much higher than the 2n = 52 predicted by chromosome painting, at 2n = 120. This is a result of the fragmentation of the results; of the 60 APCFs generated, 20 of them were smaller than the smallest mouse chromosome.

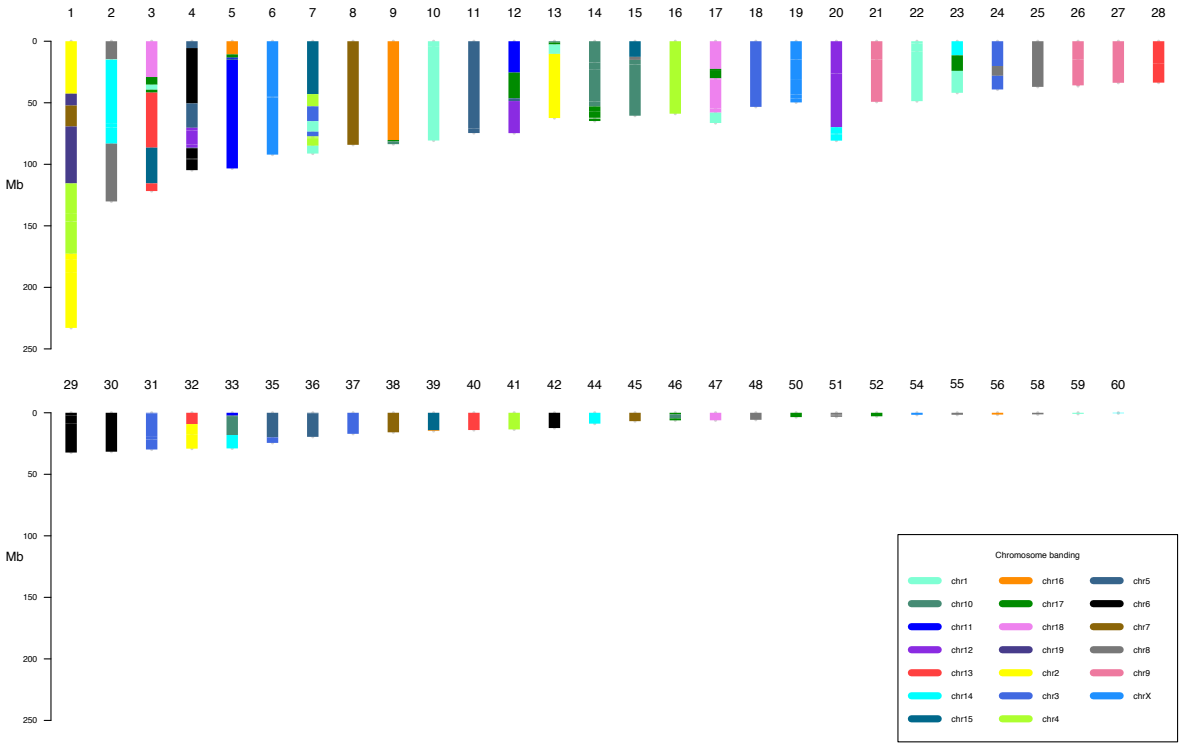


Figure 20 - Ideogram of APCFs for Muroidea ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from mouse chromosomes

There are far fewer similarities between the chromosome painting predicted ancestral karyotypes and the computational reconstruction in Muroidea than there were in the Muridae reconstruction. The associations shared between the predictions are MMU 10/17, 4, 3 and 8.

There are a great number of ways in which the predictions differ. Some of this difference will be due to the overly fragmented prediction produced by DESCHRAMBER, and some of the difference are likely due to the lack of resolution in chromosome painting, which will not allow for finding smaller homologies or rearrangements, some of the differences are as follows. Mouse chromosome 9 is maintained within one predicted chromosome in Figure 18, however is split between

APCF21, APCF26 and APCF27 in the computational reconstruction Figure 20. APCF23 in the computational reconstruction has the association MMU 14/17/1 which is potentially a fusion between the MMU 14 and MMU 17/1 associations seen in Figure 18. The association MMU 17/1/17 seen in Figure 18, which was also present in the Muridae ancestor prediction Figure 15, is present within APCF3 of the computational reconstruction Figure 20, however this is also combined with the associations MMU 18 and MMU 13/15/13, which are represented separately in the chromosome painting prediction.

Comparison to Cytogenetic Studies - Rodentia

A predicted Rodentia ancestral karyotype with a diploid number of 2n = 46 was produced [152] using comparative squirrel genome maps [153] and alignments of mouse, rat and human as seen in *Figure 21* [B]. Later another Rodentia ancestral karyotype was predicted with a diploid number of 2n = 50 [47] as seen in *Figure 21* [A]. This reconstruction was based on chromosome painting studies using *Castor fiber* (European beaver), *Sicista betulina* (birch mouse), *Pedetes capensis* (springhare) and already published studies from various squirrel species [153–156]. There are two differences between these two predictions. In the 2n = 46 prediction there is an association of HSA 20/15/14 in one predicted chromosome, whereas in the 2n = 50 prediction HSA 20 is a separate chromosome and the HSA association 15/14 is a separate chromosome. The second difference is the HSA 8/4/8/12/22 association seen in the 2n = 46 prediction, which is also split into two separate chromosomes in the 2n = 50 prediction, with the associations HSA 12/22 and HSA 8/4/8.

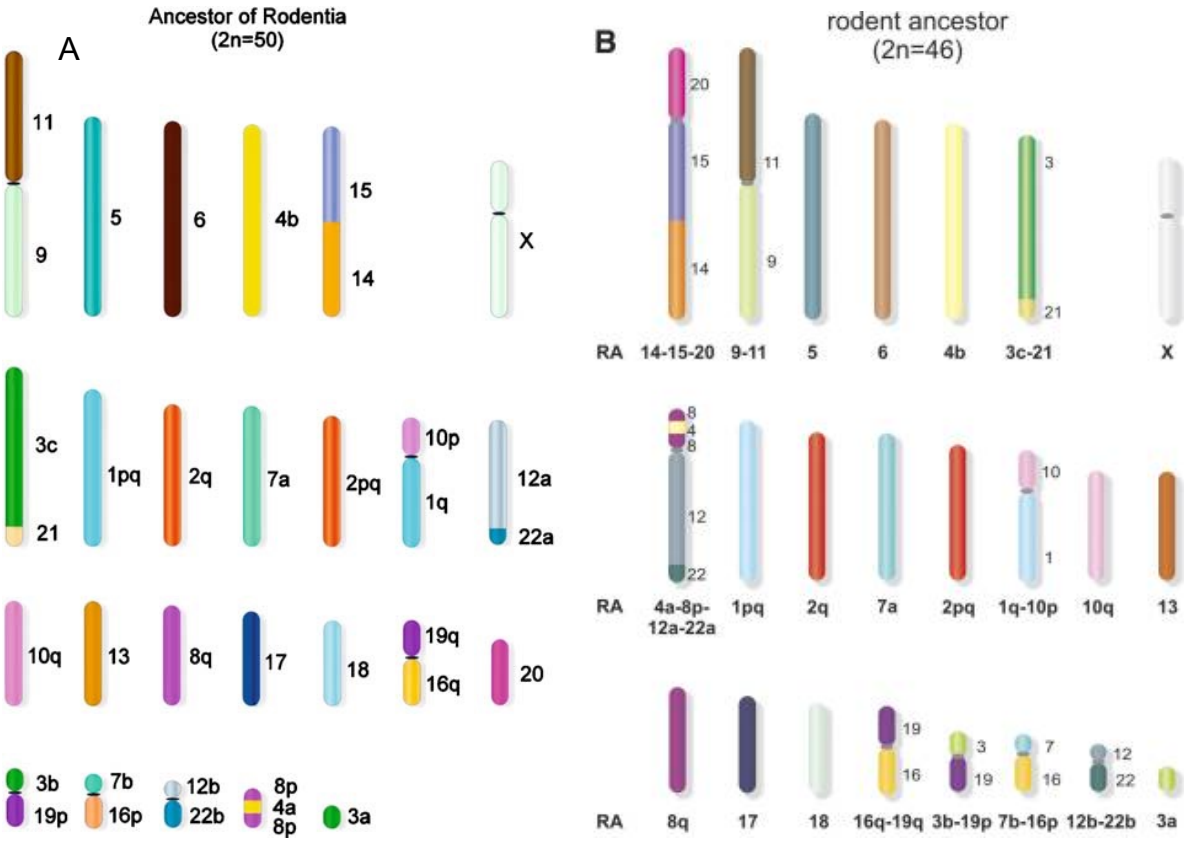


Figure 21 – [A] Ancestral karyotype of the Rodentia ancestor from cross-species chromosome painting results. Different numbers correspond to homologies of individual conserved segments in human chromosomes (HSA) [47] [B] Ancestral karyotype of the Rodentia ancestor from cross-species chromosome painting results from comparative squirrel genome maps, and on alignments of mouse, rat and human genome sequences. Different numbers correspond to homologies of individual conserved segments in human chromosomes (HSA) [152]

The reason for the differences between the two predictions could be down to the selection of species used in each. Both studies will be heavily influenced by the squirrel lineage, which has been found to be have highly conserved genome organisation [155], as multiple species in this group are used in both predictions. However, in the 2n = 46 prediction the other rodent species investigated were restricted to the Muridae family, whereas the 2n = 50 prediction used species representing a wider variety of rodent families, potentially making it more representative of the order as a whole.

The APCFs of the Rodentia ancestral prediction produced by DESCHRAMBLER are summarised in Figure 22 as ideograms made up of the homologous regions with the

human genome. The number of APCFs predicted by this method is much higher than the 2n = 50 or 2n = 46 predicted by chromosome painting, at 2n = 108. As with all of the other predictions previously mentioned, this is likely as a result of the fragmentation of the results.

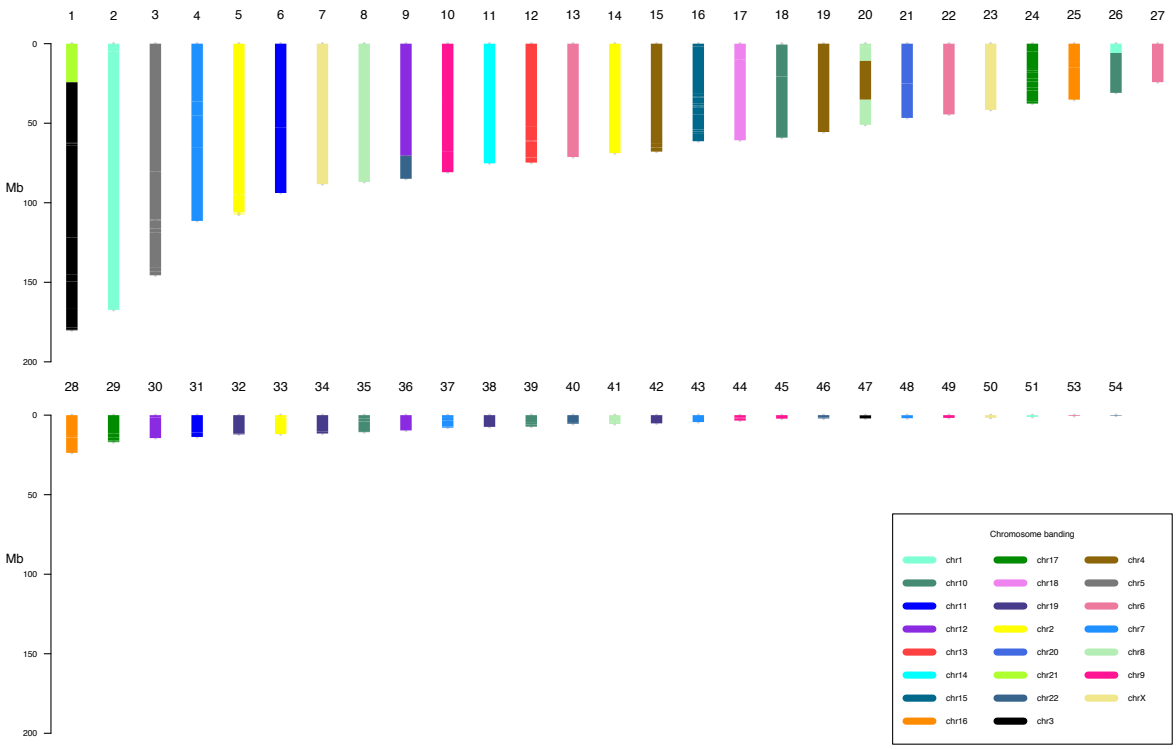


Figure 22 - Ideogram of APCFs for Rodentia ancestor produced by DESCHRAMBLER. Coloured blocks are indicative of syntenic fragments from human chromosomes

Despite the fragmentation of APCFs, there were still a number of associations found which are also found in the chromosome painting predictions, including the associations: HSA 21/3, 5, 12/22, 8/4/8, 20, and 1/10. The HSA 20 and 8/4/8 are currently more similar to the 2n = 50 prediction with them being separate chromosomes, however due to the fragmentation of this reconstruction it is not currently clear whether this is a result of over fragmentation, or our results simply incline more towards this prediction rather than the 2n = 48 prediction.

Identification of Chromosomal Rearrangements

The Genome Rearrangements In Man and Mouse (GRIMM) [129] algorithm was used to determine the number and type of chromosomal rearrangements present at each stage leading from predicted Rodentia ancestor to *Mus musculus.*

Table 5 - Number and type of rearrangements between reconstructed rodent ancestors

Ancestor	Evolutionary Time (MYA)	Number of Rearrangements				
		Inversion	Fusion	Fission	Translocation	Total
Muridae → Mouse	20.9	15	28	2	7	52
Eumuroidea → Muridae	11.8	5	0	9	12	26
Muroidea → Eumuroidea	12.3	17	20	3	27	67
Myodonta → Muroidea	10	11	15	3	28	57
Mouse Clade → Myodonta	14.9	17	0	24	23	64
Mouse Clade + Ctenohystricia → Mouse Clade	1.1	6	3	1	8	18
Rodentia → Ctenohystricia + Mouse Clade	2	4	4	1	10	19

A total of 303 rearrangements were identified using GRIMM, across 7 different species/ancestor intervals as seen in *Table 5*. Translocations were the most numerous rearrangements found, with 115. Fissions were the most infrequent at 43.

The 1.1 MYA between the ancestor of the mouse lineage and Ctenohystricia and the mouse lineage ancestor resulted in the lowest number of rearrangements with a total of 18. This supports the observation from the Evolution Highway view that large sections of these two ancestral predictions have structural homology. The 12.3 MYA between Muroidea and Eumuroidea had the highest number of rearrangements with a total of 67. Muroidea had one of the largest number of APCFs, with 55 APCFs, which is likely why there are a large number of fusions leading to Eumuroidea which was reconstructed to 34 APCFs. This same pattern of a high number of fusions is seen between Myodonta and Muroidea which have a reduction of 18 APCFs between them,

and between Muridae and Mouse with a reduction of 28 APCFs. This is suggestive that not all of the fusions observed are 'true' fusions, and that they are more likely to be due to the over fragmentation of the reconstructed ancestral predictions.

Of all the chromosomal rearrangements detected, inversions are the most likely to be accurate, due to them not being involved in a change in chromosome number, and therefore not affected by the overly fragmented reconstructions. In terms of evolution, inversions are of particular interest due to their link to recombination suppression and speciation [157,158].

Rates of Chromosomal Rearrangements

Rates of rearrangements were also calculated in the form of number of rearrangements per million years, as seen in *Table 6*.

Table 6 - Rate of rearrangements by rearrangement type between reconstructed ancestors

Ancestor	Evolutionary Time (MYA)	Rearrangement rate per MYA					
		Inv	Fus	Fiss	Trans	All	FDR corrected *p*-value
Muridae → Mouse	20.9	0.7	1.3	0.1	0.3	2.5	0.169
Eumuroidea → Muridae	11.8	0.4	0.0	0.8	1.0	2.2	0.169
Muroidea → Eumuroidea	12.3	1.4	1.6	0.2	2.2	5.5	0.654
Myodonta → Muroidea	10	1.1	1.5	0.3	2.8	5.7	0.654
Mouse Clade → Myodonta	14.9	1.1	0.0	1.6	1.5	4.3	0.377
Mouse Clade + Ctenohystricia → Mouse Clade	1.1	5.5	2.7	0.9	7.3	16.4	0.014
Rodentia → Ctenohystricia + Mouse Clade	2	2.0	2.0	0.5	5.0	9.5	0.300

The average rearrangement rate was 6.6 per MYA. There is a general trend of a reduction of rearrangement rates within the mouse lineage as compared to the Rodentia ancestor and the ancestor between the mouse lineage and Ctenohystricia

for all chromosomal rearrangement types. The greatest rate of rearrangement was 16.4 per MYA between the point at which the Ctenohystricia diverged and the ancestor for the mouse related lineage, which was significantly higher than the average rate of 6.6 per MYA (FDR corrected *p*-value of 0.014).

Figure 23 shows a phylogenetic tree of Rodentia with inversion rates between ancestors. Average rate of inversion was 1.7 per MYA. Inversion rates between Eumuroidea and Muridae, and Muridae and Mouse were lower than average. The remaining inversion rates were higher than the average, with the greatest rate of rearrangement being seen between the ancestor of the mouse lineage and Ctenohystricia, with an inversion rate of 5.5 per MYA, which was significantly higher than the average rate of inversion at 1.7 per MYA (FDR corrected *p*-value of 0.008).

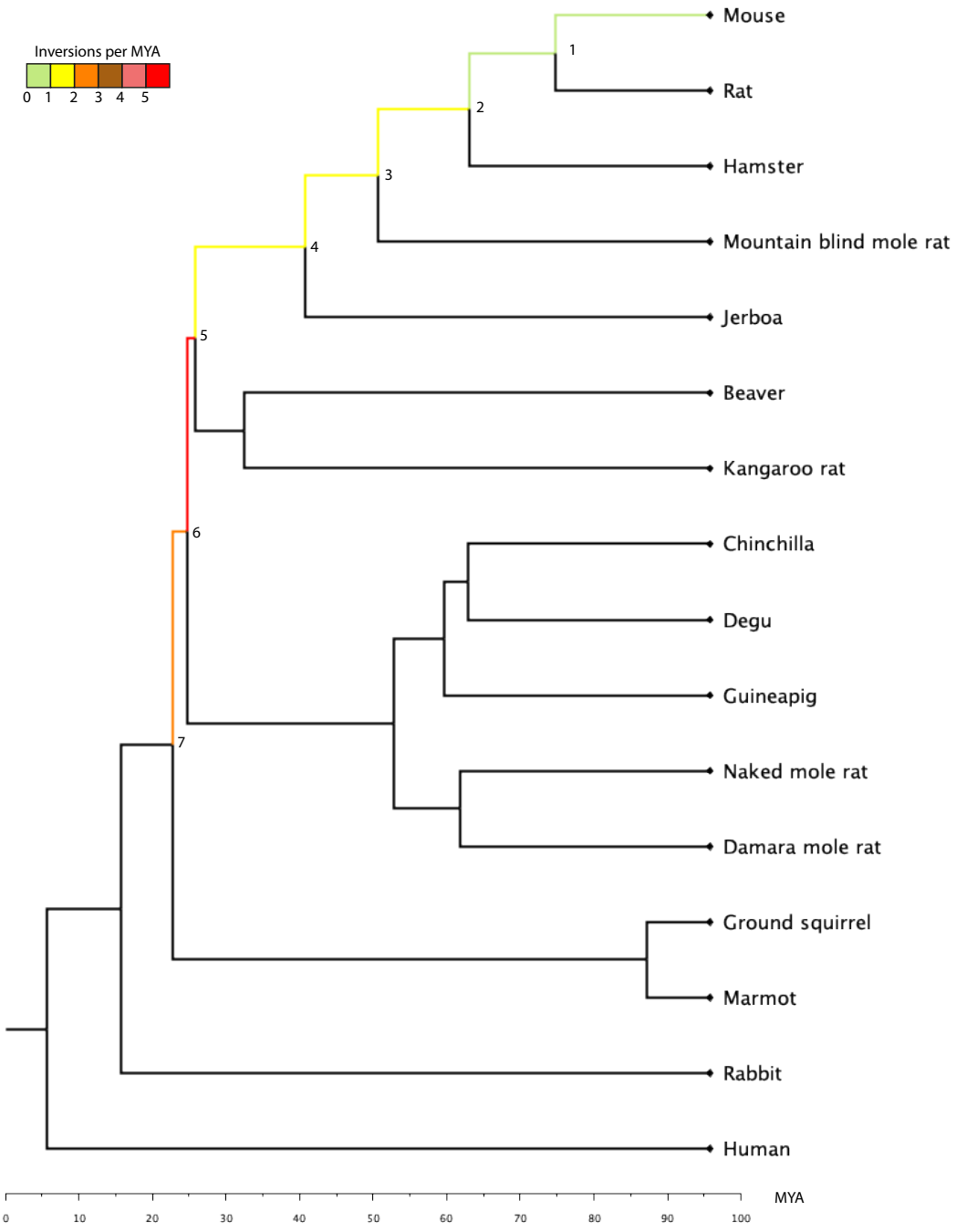


Figure 23 - Phylogenetic tree of rodent species showing rate of chromosomal inversions between reconstructed ancestors. Numbered nodes represent the following ancestors: 1 – Muridae, 2 – Eumuroidea, 3 – Muroidea, 4 – Myodonta, 5 – Mouse lineage, 6 – Mouse lineage + Ctenohystricia, 7 – Rodentia

Discussion

The predicted ancestral reconstructions produced as part of this research are by no means a complete picture of the evolutionary history of the order Rodentia, but instead

serve as a starting block to rebuilding their evolutionary history and ancestral karyotypes. The reconstructions produced are consistently fragmented, and have a diploid number far higher, sometimes double, that of similar reconstructions produced using reciprocal chromosome painting studies.

There are a few approaches that could be adopted to attempt to produce better reconstructions in the future. The first is to attempt the reconstruction again when a greater variety of high-quality assemblies become available. This reconstruction was carried out with a mind to represent as many rodent families as possible, however 14 species in an order as massive and varied as Rodentia is arguably not necessarily representative. With the launch of the Earth BioGenome project [159] it is likely that there will soon be an abundance of genomes to work with which may lead to a greater insight into their evolutionary history.

DESCHRAMBLER was run at a syntenic fragment resolution of 300 Kbp, however it is possible to run the algorithm at different syntenic fragment resolutions. In the reconstruction of the Avian ancestor, the Neognathae ancestor was reconstructed using different syntenic resolutions of 100 Kbp, 300 Kbp, and 500 Kbp, with the 100 Kbp resolution being found to produce the lowest number of APCFs, but also provide the greatest genome coverage [70]. A similar approach could be adopted with the Rodentia ancestor, see if another resolution would find the balance between identifying the finer scale rearrangements without introducing too much fragmentation in the reconstruction [69].

The development of a tool as future work to aid in the processing of manually merging the APCFs produced by the DESCHRAMBLER algorithm would have the benefit of avoiding errors being introduced into the reconstruction. At present, any combining of APCFs must be carried out by editing the raw data manually, leaving plenty of scope for human error which could range in severity from a small transposition resulting in a fragment being displaced by a few base pairs, to a section of APCF being mislabelled completely. This opportunity for human error is exacerbated by there being a number of raw data files that will need changing reciprocally for every merging of an APCF. Accurate information here is of absolute importance, not only for producing reconstructions, but also for downstream analysis of gene expression, as these start and end positions are used to intersect with the location of genes.

The cytogenetic predictions used as a comparison in this study [47,125,126,128] could also be used as guidelines to further patch together the obtained APCFs. There are a number of intermediate ancestors currently lacking an ancestral prediction, however the abundance of chromosome painting data within rodent species [126] would allow for cytogenetic based predictions of these intermediates. Although cytogenetic reconstructions could aid in the merging of APCFs, the lower resolution of chromosome painting does not allow for the identification of intrachromosome rearrangements, such as inversions. As such the cytogenetic predictions would only be used as a guide to reducing fragmentation of computational approaches, which would be relied upon to make predictions to rearrangements.

The average rate of rearrangement was found to be 6.6 rearrangements per MYA, which is greater than that found in other similar constructions from Avian ancestor to

zebra finch (2.01 per MYA) [70] and Eutherian ancestor to human (1.8 per MYA) [69]. This supports previous studies which show that rate of rearrangement is varied between lineages, with the rodent lineage having one of the highest rearrangement rates [160]. The number of interchromosomal rearrangements between the chicken and mammalian ancestor across 500 million years of evolution, only marginally exceeds the number found in 87 million years across the mouse lineage [161].

Gene Expression Analysis

RNA-Seq Alignment and Gene Counts

RNA-Seq data for liver was sourced from ENA for 7 species: *Mus musculus, Rattus norvegicus, Nannospalax galili, Fukomys damarensis, Cavia porcellus, Heterocephalus glaber* and *Oryctolagus cuniculus.* RNA-Seq data for testes was sourced from ENA for 6 species: *Mus musculus, Rattus norvegicus, Fukomys damarensis, Cavia porcellus, Heterocephalus glaber* and *Oryctolagus cuniculus*. The dataset for *Oryctolagus cuniculus* was obtained from a time-series study, to prevent the developmental stage of the animal from being a confounding variable in this case, only samples from adult individuals were used in this study. Liver tissue was selected due to the level of homogeneity of cell types within the tissue. Testes tissue was selected due to this tissue being able to reflect recombination suppression during the crossing over stage during meiosis.

Alignment of the data was carried out using STAR aligner [133], and counts produced by HTSeq-count [134]. Due to a number of the RNA-Seq datasets being from a single ended protocol, and a number of them being from a paired end protocol, each individual FASTQ file from those datasets from a paired end protocol were treated as though they were single end, to maintain consistency between all samples. To ensure that this process was not losing a significant amount of data, STAR aligner and HTSeq were run on the *Mus musculus* liver with different sets of parameters. In the first instance, STAR aligner was set for single end data, and HTSeq was run on the basis of a non strand specific protocol, seen in the samples denoted with a '_1'. Secondly, STAR aligner was set for paired end data, and HTSeq was run on the basis of a non

strand specific protocol, seen in the samples denoted with a '_2'. Finally, STAR aligner was set for paired end data, and HTSeq was run on the basis of a strand specific protocol, seen in the samples denoted with a '_3'.

Across the three samples tested with differing parameters, the option treating the samples as single end data resulted in, on average, a 2.45% increase in total number of gene counts compared to treating it as paired end data. In contrast to this, the option treating the sample as single end data resulted in, on average, a 2.14% decrease in the total number of unique reads. The increased number of gene counts read were found in the 'no feature', 'ambiguous', and 'alignment not unique' categories. Of the three sets of parameters tested, paired end alignment and strand specific count consistently produced the highest number of unique reads, single end alignment non strand specific count consistently produced the highest number of 'alignment not unique' reads, and paired end alignment non strand specific count consistently produced the highest number of ambiguous reads.

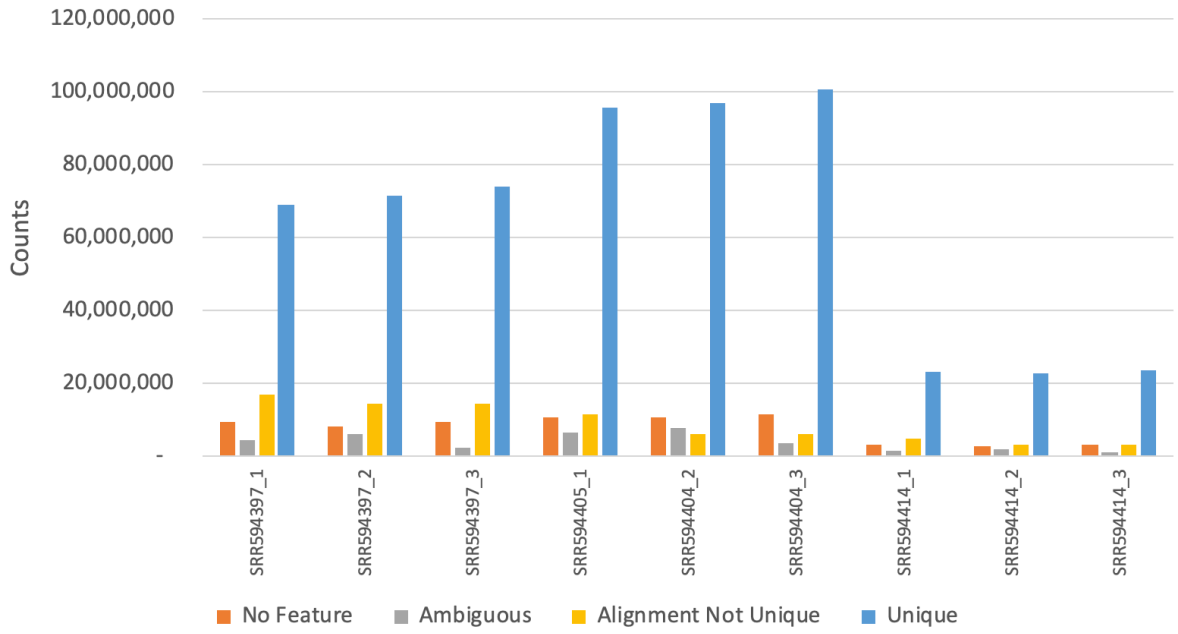


Figure 24 - Gene counts for *Mus musculus* liver RNA-Seq data under different STAR and HTSeq parameters. Samples denoted with '_1' were treated as single end, with a non strand specific protocol. Samples denoted with '_2' were treated as paired end, with a non strand specific protocol. Samples denoted with '_3' were treated as paired end, with a strand specific protocol.

Due to the low percentage of difference in read counts resulting from the different parameters, all RNA-Seq data sets were aligned with STAR as single end samples, and counted with HTSeq as a non strand specific protocol to maintain consistency across all samples. The full results of this can be seen in *Appendix Table 8.*

Of the aligned and counted RNA-Seq data, *Heterocephalus glaber* was the most successfully aligned and counted, with an average of 91% of input reads successfully aligned, and an average of 70% of these successfully aligned reads being counted as unique reads, resulting in 643,884,709 uniquely counted reads. *Oryctolagus cuniculus* was the least successfully aligned and counted, with an average of 71% of input reads being successfully aligned, and an average of 68% of these successfully aligned reads being counted as unique reads, resulting in 97,893,222 uniquely counted reads. Uniquely counted reads for the remaining species were: 739,241,003 for *Mus*

musculus, 617,327,711 for *Rattus norvegicus*, 255,488,050 for *Fukomys damarensis,* and 640,945,972 for *Cavia porcellus* respectively. A higher percentage of uniquely counted reads was observed in liver samples than in testes samples in all species with the exception of *Mus musculus.*

Filtering for Orthologous Genes

Gene orthologues for *Mus musculus, Rattus norvegicus, Nannospalax galili, Cavia porcellus, Fukomys damarensis, Heterocephalus glaber* and *Oryctolagus cuniculus* were downloaded from Ensembl release 97 [137]. Due to the reconstructions being generated by a number of different rodent species, orthologous genes need to be used to ensure that the same gene(s) is being investigated across all species, rather than unique genes from each species. Resultant orthologues were filtered to use only one-to-one orthologues. 9,883 one-to-one orthologues were identified in total, 7,411 of which were found in all of the species selected. The unique Gene ID for *Mus musculus* was retained for each orthologue, to allow for later mapping against predicted rearrangements, due to having gene location data for *Mus musculus.* One-to-one orthologues are required to accurately see any potential changes in expression levels in genes across evolutionary time. Using one-to-many, many-to-one or many-to-many orthologues could result in not comparing the same genes in each species.

Principal Component Analysis

The gene counts were normalised using DESeq2 [135] and Principal Component Analysis (PCA) carried out, *Figure 25*. It was found that when samples were from different tissues in different species, the samples cluster preferentially by tissue over species, with liver samples represented by circles clustering together on the left, and

testes sample represented by triangles clustering together on the right. This supports results found in previous RNA-Seq studies using different tissues from multiple species [162,163] which found that conservation of organ physiology led to the clustering by tissue rather than species. As the aim of this study is to investigate the role of chromosomal rearrangements in the speciation of rodents, the RNA-Seq data was therefore separated into two datasets, one for liver and one for testes, to ensure that gene expression change between species is the focus of analysis.

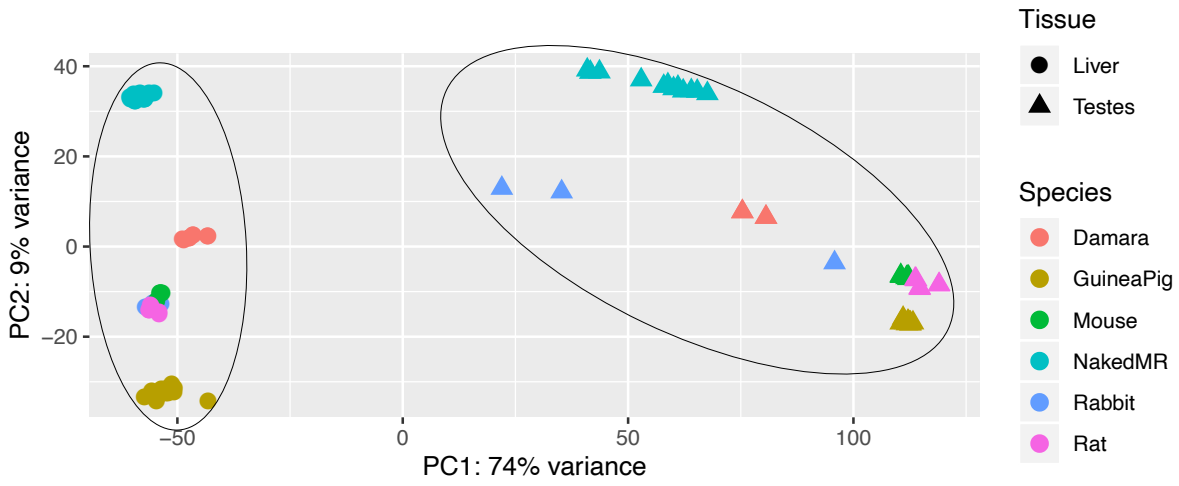


Figure 25 - Principal component analysis of gene expression levels in liver and testes tissues of 5 rodent species and 1 outgroup species

Within the tissue clusters, species then clustered together as represented by their colours, with the exception of rabbit within the testes cluster. This is likely due to the rabbit testes samples having both the least successful alignment, with 67% of input reads being successfully aligned, and also the lowest number of these successfully aligned reads being classified as unique, with an average of 62%.

Gene Expression Correlation

Mean expression data was calculated for the liver and testes of each species, and mean expression data across all species for each tissue was also calculated. Expression data was then subset by species. The correlation of gene expression was then calculated pairwise between species, using Spearman's rank correlation. Pairwise correlation values for gene expression in all genes in both liver and testes can be seen in *Figure 26*.

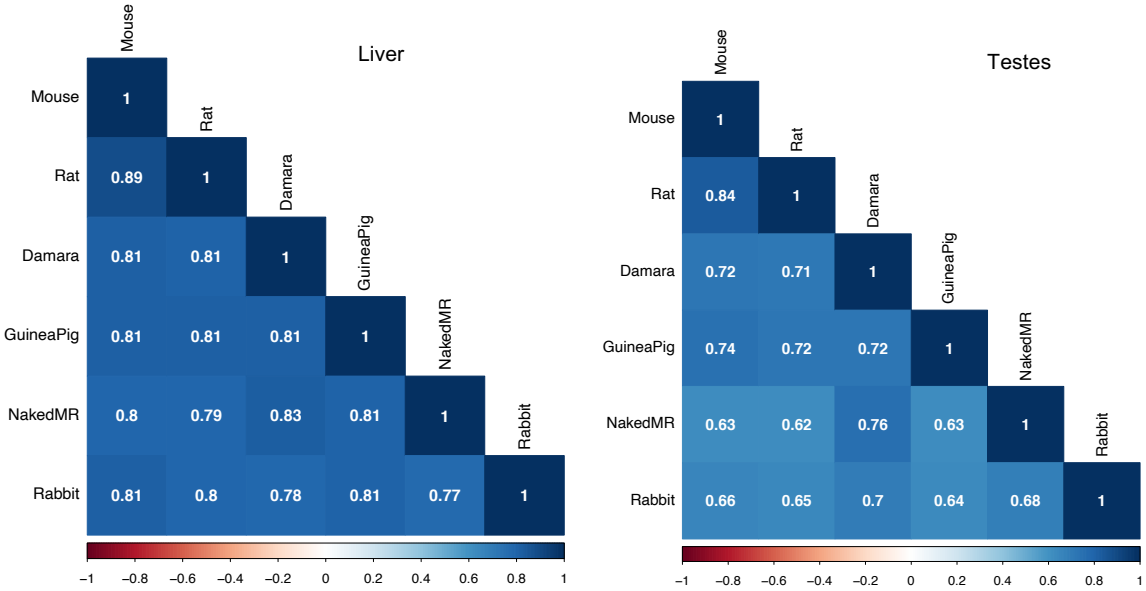


Figure 26 - Correlation plot of gene counts in liver and testes in rodent species using Spearman's Rank correlation

Pairwise correlations in both liver and testes tissues were higher in species that were more closely related to each other, than those that are more distantly related. In both liver and testes tissues mouse and rat have the highest correlation, and both species are in the family Muridae. In contrast to this in the liver tissue the lowest pairwise correlation is between mouse from Order Rodentia, and rabbit from Order Lagomorpha.

Pairwise correlations between species in liver tissue were ubiquitously higher in all species pairings when compared to pairwise correlations in testes tissue. The highest correlation in liver tissue being 0.89 between mouse and rat, compared to the highest correlation in testes tissue of 0.84 between mouse and rat, 0.05 lower. The lowest correlation in liver tissue was 0.77 between rabbit and naked mole rat, whereas the lowest correlation in testes tissue was 0.62 between rat and naked mole rat, 0.15 lower.

Gene Expression in Rearrangements

To identify which of the gene orthologues used were found within the ancestral reconstructions, biomaRt [137] was used to find the location of each gene within the genome of *Mus musculus.* These locations were then compared to the output map files from DESCHRAMBLER for the Muridae ancestral reconstruction which show the SFs between either one ancestor and another, or between reference species and ancestor, in this case using the SFs between the Muridae ancestors and mouse. The intersection of these two datasets was carried out by using the BEDTools intersect function [139].

Orthologues Absent from Muridae Reconstruction

7,387 gene orthologues successfully mapped to the syntenic fragments making up the Muridae ancestor. 24 gene orthologues were not successfully mapped to the syntenic fragments, these genes can be seen in *Table 7*. 92.78% of the *Mus musculus* genome was covered in the Muridae ancestral reconstruction *Table 4*, so it would be reasonable to expect that a small proportion of orthologues would be omitted from the

reconstruction. The genes not mapped represent 0.3% of one to one orthologues used, 41.6% of those in *Table 7* are found in *Mus musculus* chromosome 17, which was one of the most fragmented chromosomes used in the reconstruction with frequent gaps found between SFs.

Table 7 - Orthologues genes missing from the Muridae ancestor reconstruction

GeneID	Gene	Chr	Start (bp)	End (bp)
ENSMUSG00000014932	Yes1	5	32611171	32687057
ENSMUSG00000014956	Ppp1cb	5	32458843	32517433
ENSMUSG00000019872	Smpdl3a	10	57794335	57811830
ENSMUSG00000019874	Fabp7	10	57784881	57788450
ENSMUSG00000021518	Ptdss1	13	66932830	66998401
ENSMUSG00000021519	Mterf3	13	66906968	66933088
ENSMUSG00000023940	Sgo1	17	53674786	53689333
ENSMUSG00000023965	Fbxl17	17	63057452	63500017
ENSMUSG00000024193	Phf1	17	26933052	26937908
ENSMUSG00000024194	Cuta	17	26933819	26939569
ENSMUSG00000024227	Pdzph1	17	58878808	58991375
ENSMUSG00000024228	Nudt12	17	58999618	59013372
ENSMUSG00000025747	Tyms	5	30058202	30073617
ENSMUSG00000025898	Cwf19l2	9	3403592	3479236
ENSMUSG00000036928	Stag3	5	138280240	138312393
ENSMUSG00000039497	Dse	10	34151393	34207715
ENSMUSG00000039508	Calhm4	10	34038784	34044310
ENSMUSG00000039531	Zup1	10	33919142	33951269
ENSMUSG00000042644	Itpr3	17	27057304	27122223
ENSMUSG00000048915	Efna5	17	62604184	62881317

ENSMUSG00000049872	Calhm5	10	34087815	34096519
ENSMUSG00000057789	Bak1	17	27019810	27029009
ENSMUSG00000067629	Syngap1	17	26941253	26972434
ENSMUSG00000071340	Trappc3l	10	34037597	34109815

Once each gene had been mapped to its respective SF, the GRIMM results showing which SFs were involved in different chromosomal rearrangements, could then be incorporated to show which orthologues were located in inversions, fusions, fissions, and translocations. It was found that 606 were located within inversions, 3,667 were found within fusions, 415 were found in fissions, and 880 were found within translocations. Inversions were the type of chromosomal rearrangement selected to be investigated further due to the increased likelihood of them having been accurately predicted.

Rearrangements in Inversions - Liver

Mean expression data was next further subset into two groups, one group representing genes which were found to be present within inversions, and another group for those genes which were not present in inversions. Pairwise correlations were then calculated between each species for genes located in inversions, and between each species for genes not present within inversions.

Pairwise correlations for inversions and non-inversions in liver tissue can be seen in *Figure 27*. 12 out of 15 correlation values were found to be lower in genes found within inversions when compared to genes not within inversions, the other 3 correlation values remained the same between conditions. Damaraland mole rat and guinea pig unanimously had lower correlations in the subset of genes within inversions.

The paired difference between these two conditions was calculated using a Wilcoxon signed rank test, resulting in a *p*-value of 0.032.

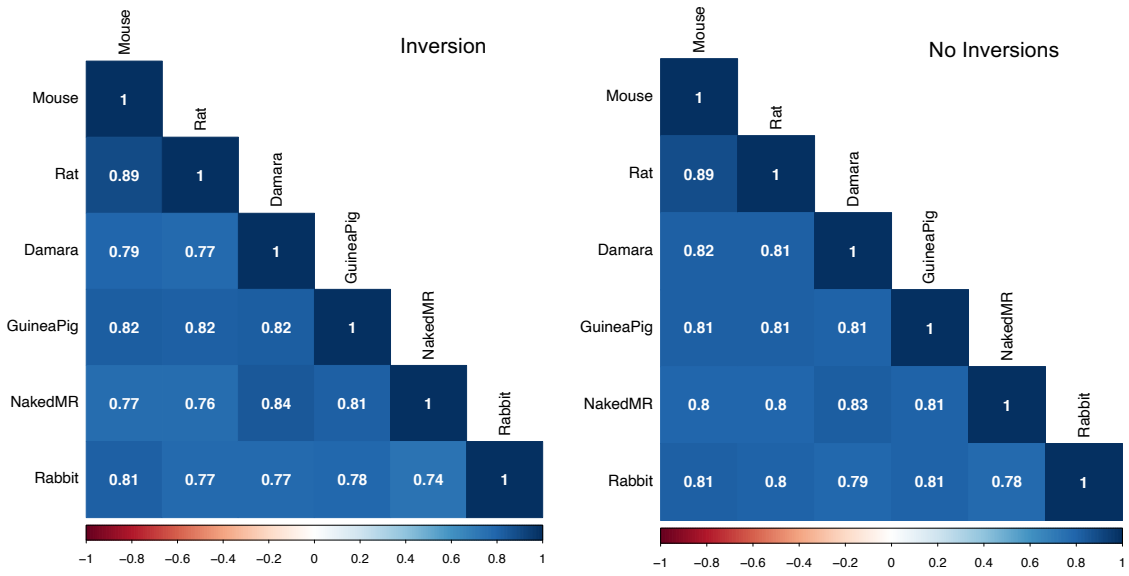


Figure 27 - Correlation plot of liver gene counts in inversions in the Muridae ancestor (left) and those not in inversions in the Muridae ancestor (right) using Spearman's Rank correlation

There is a chance that comparing correlation between these two conditions could be confounded by extremes of expression value in some genes, particularly when the number of genes not present in inversions is 10 times higher than those present in inversions. To prevent this from happening, MatchIT [140] was used to generate a subset for those genes not in inversions which, in terms of mean expression for that gene, was closer to the characteristics of the subset of genes with inversions. It uses propensity score matching to select genes not found in inversions which have comparable mean expression data across all species, to the mean expression data across all species for the subset of genes with inversions. Thereby reducing the chance that the result is overly affected by extremes of expression levels.

Pairwise correlations for matched inversions and non-inversions in liver tissue can be seen in *Figure 28*. 5 out of 15 correlation values were found to be lower in genes found within inversions when compared to genes not within inversions. 6 out of 15 correlation values were found to be higher in genes found within inversions when compared to genes not within inversions. The remaining 4 correlation values remained the same between both conditions.

The paired difference between these two conditions was calculated using a Wilcoxon signed rank test, resulting in a *p*-value of 0.445.

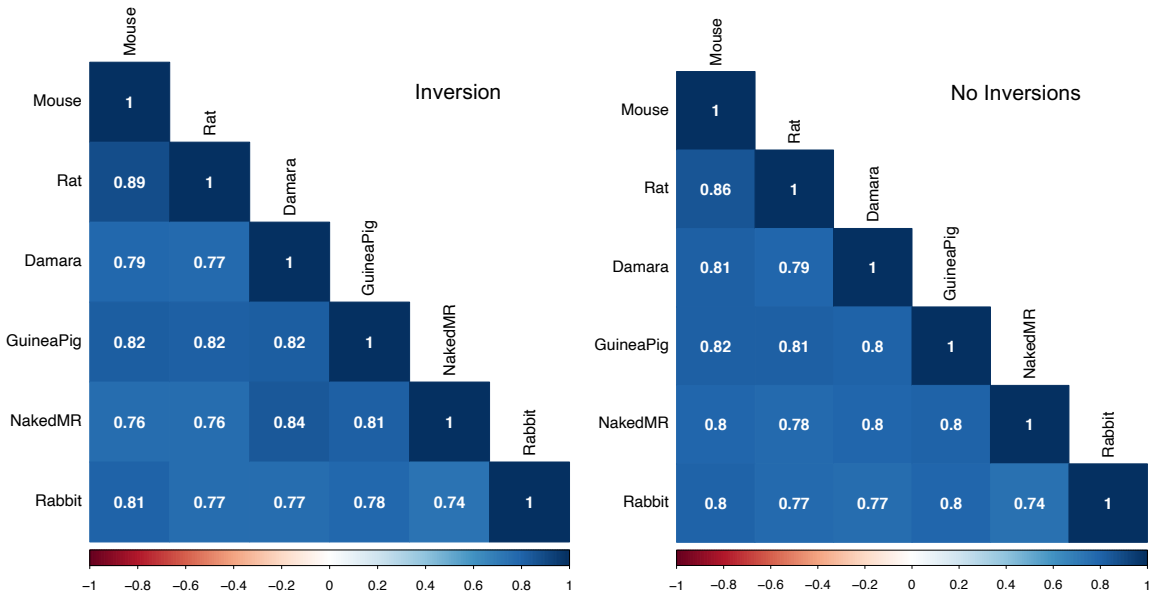


Figure 28 - Correlation plot of matched liver gene counts in inversions in the Muridae ancestor (left) and those not in inversions in the Muridae ancestor (right) using Spearman's Rank correlation

Rearrangements in Inversions - Testes

Pairwise correlations for inversions and non-inversions in testes tissue can be seen in *Figure 29*. 7 out of 15 correlation values were found to be lower in genes found within inversions when compared to genes not within inversions. Only rabbit had lower correlation scores in all pairwise species comparisons. In contrast to the correlations in liver tissue, 7 out of 15 correlation values were found to be higher in the subset of inversions. One correlation value remained the same between conditions.

The paired difference between these two conditions was calculated using a Wilcoxon signed rank test, resulting in a *p*-value of 0.533.

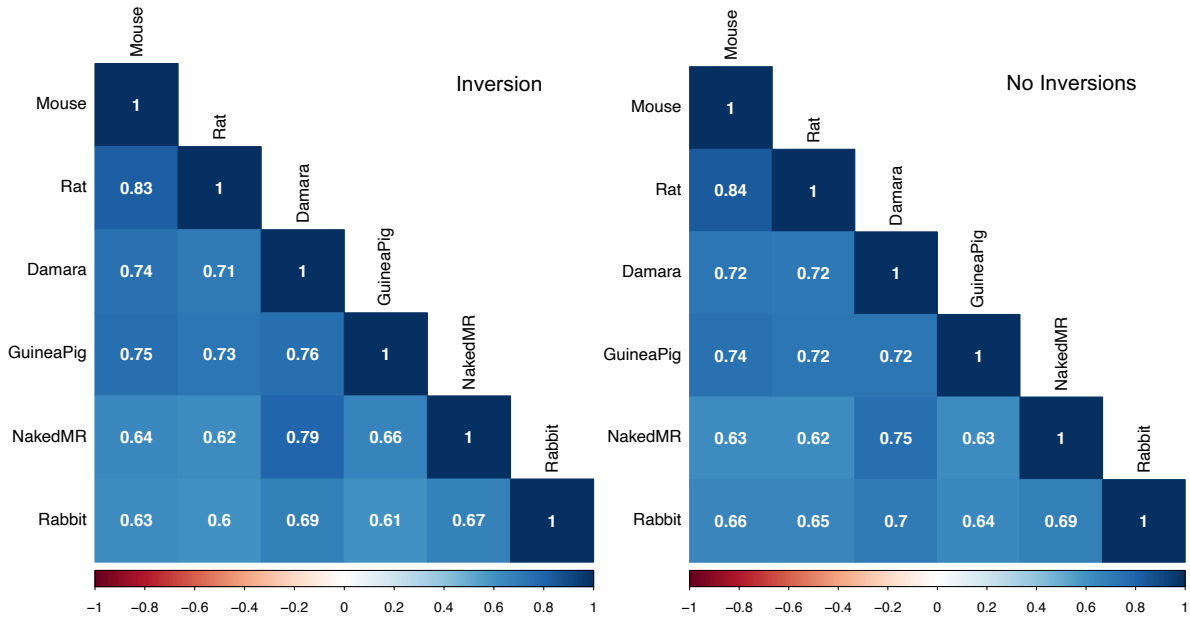


Figure 29 - Correlation plot of testes gene counts in inversions (left) and those not in inversions (right) using Spearman's Rank correlation

Testes gene expression was also subset using propensity score matching, as previously described for the liver expression data. Pairwise correlations for matched

inversions and non-inversions in testes tissue can be seen in *Figure 30*. 9 out of 15 correlation values were found to be lower in genes found within inversions when compared to genes not within inversions. 5 out of 15 correlation values were found to be higher in genes found within inversions when compared to genes not within inversions. 1 correlation value remained the same between conditions.

The paired difference between these two conditions was calculated using a Wilcoxon signed rank test, resulting in a *p*-value of 0.060.

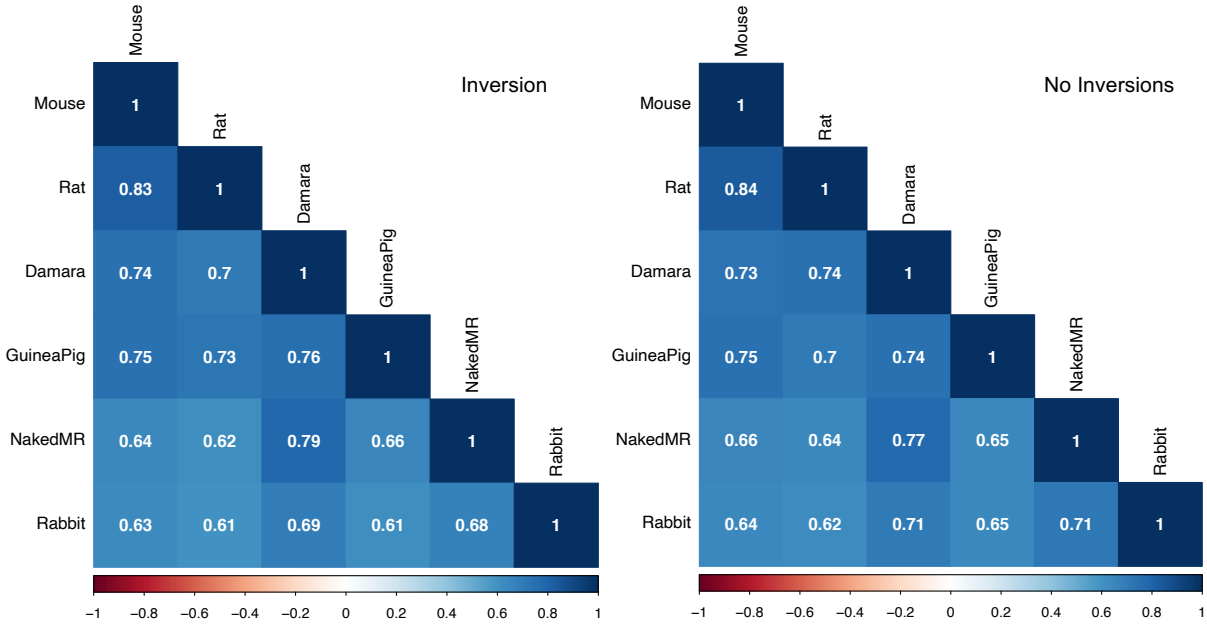


Figure 30 - Correlation plot of matched testes gene counts in inversions (left) and those not in inversions (right) using Spearman's Rank correlation

Gene Ontology in Inversions

To find if there are categories of genes associated with inversions between ancestor and species GO enrichment analysis, in the form of a statistical overrepresentation test, was carried out in Panther [141]. The genes mapped to inversions between *Mus musculus* and the predicated Muridae ancestral reconstruction were used, and the entire gene set for *Mus musculus* was used as the reference dataset. Terms with a *p* value of < 0.05 and a false discovery rate (FDR) of < 5% were considered to be significantly enriched. Classes of GO enrichment investigated were biological processes, cellular components, and molecular function, the results of which are summarised in *Figure 31*.

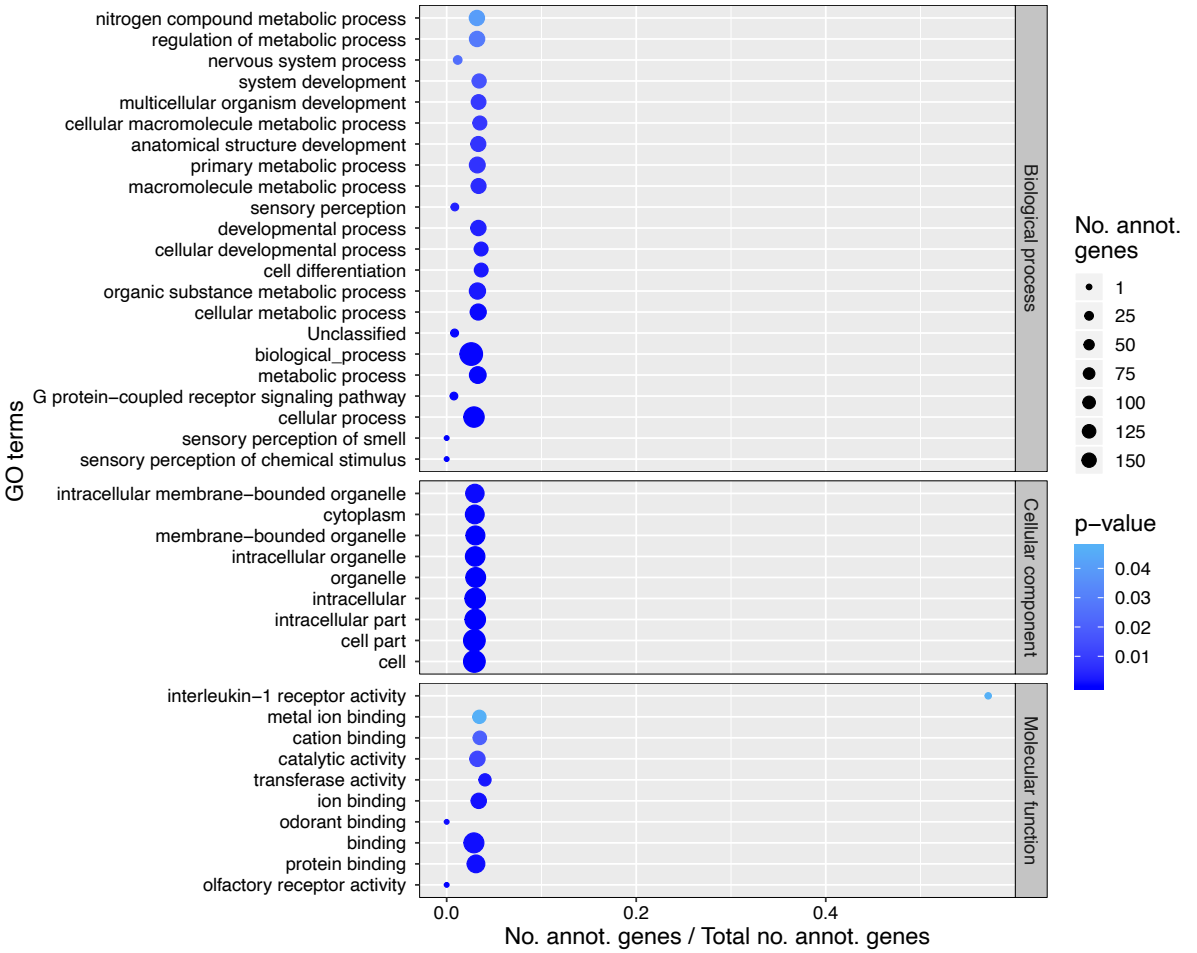


Figure 31 - GO terms enriched in the inversions between Muridae and Mus musculus with a *p*-value < 0.05 and FDR < 5%

In the GO enrichment class for biological process, genes enriched for *cell differentiation, cellular macromolecule metabolic process,* and *system development* were found amongst others. Genes related to *cellular developmental process* (n = 137) were found to have the highest fold enrichment against the background gene set (*p*-value = 0.00239). Some GO processes were found to have a fold enrichment lower in the gene set with inversions than would be expected against the background gene set. These processes are: *G protein-coupled receptor signalling pathway* (n = 14), *sensory perception of smell* (n = 0), *sensory perception of chemical stimulus* (n = 0), *sensory perception* (n = 14), and *nervous system process* (n = 24).

In the GO enrichment class for cellular component, genes enriched for *intracellular membrane-bounded organelle,* and *cytoplasm* were found. Genes related to *membrane-bounded organelle* (n = 332) were found to have the highest fold enrichment against the background gene set (*p*-value = 0.0000117).

In the GO enrichment class for molecular function, genes enriched for *interleukin-1 receptor activity, transferase activity*, and *metal ion binding* were found amongst others. Genes related to *interleukin-1 receptor activity* (n = 4) were found to have the highest fold enrichment against the background gene set (*p*-value = 0.0468). Some GO processes were found to have a fold enrichment lower in the gene set with inversions than would be expected against the background gene set. These processes are: *olfactory receptor activity* (n = 0) and *odorant binding* (n = 0).

Discussion

Gene expression is important to investigate when considering the implications of chromosomal rearrangements in speciation. Normal development of an individual relies on gene expression pathways, gene order and correct gene expression levels. Rearrangements which interrupt gene expression pathways or modify gene dosage can result in a differing levels of gene expression, contributing to the variety of unique traits seen between different species seen on the Earth today.

Due to the limited amount of publicly available RNA-Seq data which fitted the criteria of this study, only gene expression correlation in rearrangements for the Muridae and Rodentia ancestors had the potential to be investigated here. If baseline RNA-Seq data were to be generated in the future for some of the other species investigated as part of this study, then it would be possible to investigate gene expression in the intermediate ancestors produced earlier in this study. At this stage only gene expression correlation in the Muridae ancestor was investigated, due to there being one ‘step’ between mouse, which we had gene locations for, and Muridae, which we had rearrangement locations for. Generating the rearrangement locations for all of the ‘steps’ leading from the Rodentia to mouse would be a more complicated process, but is possible with the information available, and is a definite area of future work.

Inversions were the rearrangement of choice to investigate here in part due to them being the rearrangement predicted with the greatest accuracy during the reconstruction stage, but In also terms of evolution, inversions are of particular interest due to their link to recombination suppression and speciation [157,158]. If these inversions were contributing to speciation, we would expect to see that genes within

inversions will have a lower correlation between species than those genes found outside inversions. The results here suggest that there was not a significant difference in gene expression correlation between genes in inversions and those not, with a *p*-value of 0.445 for liver tissue and a *p*-value of 0.060 in testes tissue. Testes was far closer to being considered significant than liver, which is consistent with recombination suppression during meiosis [17].

The finding that there was no significant difference between gene expression in inversions compared to those not in inversions is surprising considering that inversions have been implicated in recombination suppression [157,164] and misexpression of genes [165], and a source of genetic variation [166]. Eukaryotic genomes are complex systems with many constituent parts which could be playing a role. For example topologically associating domains (TADs) are involved in the maintenance of the gene regulatory network with which they are associated [167] and have been shown to have conserved gene regulation within their boundaries [168] which are resilient to rearrangement, with EBRs tending to occur at TAD boundaries [168,169]. Synteny blocks have previously been shown to be enriched for evolutionary conserved sequences [170], whereas regions surrounding EBRs are gene rich regions linked to genes useful for adaptation [171]. This goes to show that presence of a gene within an inversion is not the full story, and that other factors such as location within the inversion, or distance from EBRs, or the effect of chromatin interactions and the wider genomic architecture might be responsible for changes in gene expression. Further studies into chromatin interactions/disruption around chromosomal rearrangements using methods such as chromatin immunoprecipitation sequencing (ChIP-seq), or

focusing primarily in genes found near EBRs may elucidate this matter further in the future.

Gene correlation expression could also be investigated for fusions, fissions, and translocations in Muridae and Rodentia, and would be an interesting avenue for further investigation. However, due to the uncertainty surrounding the accuracy of the number of these type of rearrangements, due to the over fragmentation of the ancestral reconstructions, this is an area of further enquiry which would be best pursued with the generation of more complete ancestral reconstructions, and therefore more accurate locations for chromosome rearrangements.

Conclusion

This study failed to implicate chromosomal rearrangements in speciation within the Rodentia order, however it does provide a foundation for further investigation into the evolutionary history of rodents.

Ancestral reconstructions were found to be highly fragmented, with all 7 ancestors displaying high diploid numbers with multiple unresolved APCFs. Comparing reconstructions with different basal clades allowed us to support the squirrel related lineage forms the base of the Rodentia phylogeny. Despite fragmentation the reconstructions still allowed for the successful identification of chromosomal rearrangements at each stage of the lineage, with the greatest confidence being in the accuracy of the inversions, and fusions the most artificially inflated by fragmentation. It is thought that more complete reconstructions would improve the accuracy of the remaining rearrangement types. Both number and rate of rearrangements were high compared to other mammalian lineages, which agrees with the current literature and also reflects the difficulty in the reconstruction of the karyotypes.

Gene expression was found to be lower in genes that were within inversions, than genes that were not in inversion as would be expected if linked to speciation, however these results were outside the bounds of statistical significance. The results of this work do not indicate that speciation is driven by inversions in Rodentia, however it is believed that future work on the reconstructions, and greater understanding of the implication of the wider genome architecture on gene expression, may lead to a more complete picture. There are still many avenues for future work to investigate before chromosomal speciation can be ruled out in this instance.

References

1. Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. (2011). How many species are there on earth and in the ocean? PLoS Biol.
2. Futuyma, D.J. (2009). Speciation. In Evolution (Sinauer Associates, Inc), pp. 471–498.
3. Wright, S. (1978). Modes of Speciation. Michael J. D. White W. H. Freeman and Co., San Francisco. 1978. VIII + 456 pp. illus. $27.50. Paleobiology.
4. Rieseberg, L.H. (2001). Chromosomal rearrangements and speciation. Trends Ecol. Evol., 351–358.
5. Harewood, L., and Fraser, P. (2014). The impact of chromosomal rearrangements on regulation of gene expression. Hum. Mol. Genet. *23*, 76–82.
6. Demura, M., Martin, R.M., Shozu, M., Sebastian, S., Takayama, K., Hsu, W.T., Schultz, R.A., Neely, K., Bryant, M., Mendonca, B.B., *et al.* (2007). Regional rearrangements in chromosome 15q21 cause formation of cryptic promoters for the CYP19 (aromatase) gene. Hum. Mol. Genet. *16*, 2529–2541.
7. Yunis, J.J., and Prakash, O. (1982). The origin of man: A chromosomal pictorial legacy. Science (80-.). *215*, 1525–1530.
8. IJdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T., and Wells, R.A. (2006). Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proc. Natl. Acad. Sci. *88*, 9051–9055.
9. Avarello, R., Pedicini, A., Caiulo, A., Zuffardi, O., and Fraccaro, M. (1992). Evidence for an ancestral alphoid domain on the long arm of human chromosome 2. Hum. Genet. *89*, 247–249.
10. Ventura, M., Mudge, J.M., Palumbo, V., Burn, S., Blennow, E., Pierluigi, M., Giorda, R., Zuffardi, O., Archidiacono, N., Jackson, M.S., *et al.* (2003).

Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. Genome Res. *13*, 2059–2068.

11. Giannuzzi, G., Pazienza, M., Huddleston, J., Antonacci, F., Malig, M., Vives, L., Eichler, E.E., and Ventura, M. (2013). Hominoid fission of chromosome 14/15 and the role of segmental duplications. Genome Res. *23*, 1763–1773.
12. Feuk, L. (2010). Inversion variants in the human genome: Role in disease and genome architecture. Genome Med.
13. Entesarian, M., Carlsson, B., Mansouri, M.R., Stattin, E.L., Holmberg, E., Golovleva, I., Stefansson, H., Klar, J., and Dahl, N. (2009). A chromosome 10 variant with a 12 Mb inversion [inv(10)(q11.22q21.1)] identical by descent and frequent in the Swedish population. Am. J. Med. Genet. Part A, 380–386.
14. Antonarakis, S.E., Rossiter, J.P., Young, M., Horst, J., de Moerloose, P., Sommer, S.S., Ketterling, R.P., Kazazian, H.H., Négrier, C., Vinciguerra, C., *et al.* (1995). Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. Blood *86*, 2206–2212.
15. Bagnall, R.D., Waseem, N., Green, P.M., and Giannelli, F. (2002). Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. Blood *99*, 168–174.
16. Kirkpatrick, M. (2010). How and why chromosome inversions evolve. PLoS Biol. *8*.
17. Griffiths AJF, Gelbart WM, Miller JH, et al. (1999). Chromosomal Rearrangements. In Modern Genetic Analysis (W. H. Freeman).
18. Brown, J.D., and O'Neill, R.J. (2010). Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited. Annu. Rev. Genomics Hum. Genet. *11*, 291–316.

19. Templeton, A.R. (1981). Mechanisms of Speciation - A Population Genetic Approach. Annu. Rev. Ecol. Syst. *12*, 23–48.

20. Lewis, H. (1966). Speciation in flowering plants. Science (80-.).

21. Manoukis, N.C., Powell, J.R., Touré, M.B., Sacko, A., Edillo, F.E., Coulibaly, M.B., Traoré, S.F., Taylor, C.E., and Besansky, N.J. (2008). A test of the chromosomal theory of ecotypic speciation in Anopheles gambiae. Proc. Natl. Acad. Sci. U. S. A. *105*, 2940–2945.

22. Navarro, A., and Ruiz, A. (1997). On the fertility effects of pericentric inversions. Genetics, 931–933.

23. Noor, M.A.F., Garfield, D.A., Schaeffer, S.W., and Machado, C.A. (2007). Divergence between the Drosophila pseudoobscura and D. persimilis genome sequences in relation to chromosomal inversions. Genetics *177*, 1417–1428.

24. Brown, K.M., Burk, L.M., Henagan, L.M., and Noor, M.A.F. (2004). A test of the chromosomal rearrangement model of speciation in Drosophila pseudoobscura. Evolution (N. Y). *58*, 1856–1860.

25. Nadeau, J.H., and Taylor, B.A. (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. Proc. Natl. Acad. Sci. *81*, 814–818.

26. Ohno, S. (1973). Ancient linkage groups and frozen accidents. Nature *244*, 259–262.

27. Copeland, N., Jenkins, N., Gilbert, D., Eppig, J., Maltais, L., Miller, J., Dietrich, W., Weaver, A., Lincoln, S., Steen, R., *et al.* (2006). A genetic linkage map of the mouse: current applications and future prospects. Science (80-.). *262*, 57–66.

28. Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., and Haussler, D. (2003).

Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. *100*, 11484–11489.

29. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.
30. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.
31. Pevzner, P., and Tesler, G. (2003). Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res. *13*, 37–45.
32. Pevzner, P., and Tesler, G. (2003). Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. Proc. Natl. Acad. Sci. *100*, 7672–7677.
33. Becker, T.S., and Lenhard, B. (2007). The random versus fragile breakage models of chromosome evolution: A matter of resolution. Mol. Genet. Genomics.
34. Sandelin, A., Bailey, P., Bruce, S., Engström, P.G., Klos, J.M., Wasserman, W.W., Ericson, J., and Lenhard, B. (2004). Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics *5*.
35. Ruiz-Herrera, A., García, F., Giulotto, E., Attolini, C., Egozcue, J., Ponsà, M., and Garcia, M. (2005). Evolutionary breakpoints are co-localized with fragile sites and intrachromosomal telomeric sequences in primates. Cytogenet. Genome Res. *108*, 234–247.
36. Ruiz-Herrera, A., and Robinson, T.J. (2007). Chromosomal instability in

Afrotheria: Fragile sites, evolutionary breakpoints and phylogenetic inference from genome sequence assemblies. BMC Evol. Biol. *7*.

37. Ruiz-Herrera, A., Castresana, J., and Robinson, T.J. (2006). Is mammalian chromosomal evolution driven by regions of genome fragility? Genome Biol. *7*, 115.
38. Corbett-Detig, R.B. (2016). Selection on inversion breakpoints favors proximity to pairing sensitive sites in Drosophila melanogaster. Genetics *204*, 259–265.
39. Peng, Q., Pevzner, P.A., and Tesler, G. (2006). The fragile breakage versus random breakage models of chromosome evolution. PLoS Comput. Biol. *2*.
40. Lemaitre, C., Zaghloul, L., Sagot, M.F., Gautier, C., Arneodo, A., Tannier, E., and Audit, B. (2009). Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation. BMC Genomics *10*.
41. Kemkemer, C., Kohn, M., Cooper, D.N., Froenicke, L., Högel, J., Hameister, H., and Kehrer-Sawatzki, H. (2009). Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. BMC Evol. Biol. *9*.
42. Farré, M., Robinson, T.J., and Ruiz-Herrera, A. (2015). An Integrative Breakage Model of genome architecture, reshuffling and evolution. BioEssays *37*, 479–488.
43. Bailey, J.A., and Eichler, E.E. (2006). Primate segmental duplications: Crucibles of evolution, diversity and disease. Nat. Rev. Genet., 552–564.
44. Farré, M., Bosch, M., López-Giráldez, F., Ponsà, M., and Ruiz-Herrera, A. (2011). Assessing the role of tandem repeats in shaping the genomic architecture of great apes. PLoS One *6*.

45. Bourque, G. (2009). Transposable elements in gene regulation and in the evolution of vertebrate genomes. Curr. Opin. Genet. Dev., 607–612.

46. Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet., 292–301.

47. Graphodatsky, A.S., Yang, F., Dobigny, G., Romanenko, S.A., Biltueva, L.S., Perelman, P.L., Beklemisheva, V.R., Alkalaeva, E.Z., Serdukova, N.A., Ferguson-Smith, M.A., *et al.* (2008). Tracking genome organization in rodents by Zoo-FISH. Chromosom. Res. *16*, 261–274.

48. Kulemzina, A.I., Trifonov, V.A., Perelman, P.L., Rubtsova, N. V., Volobuev, V., Ferguson-Smith, M.A., Stanyon, R., Yang, F., and Graphodatsky, A.S. (2009). Cross-species chromosome painting in Cetartiodactyla: Reconstructing the karyotype evolution in key phylogenetic lineages. Chromosom. Res. *17*, 419–436.

49. Murphy, W.J., Bourque, G., Tesler, G., Pevzner, P., and O'Brien, S.J. (2003). Reconstructing the genomic architecture of mammalian ancestors using multispecies comparative maps. Hum. Genomics *1*, 30–40.

50. Ferguson-Smith, M.A., and Trifonov, V. (2007). Mammalian karyotype evolution. Nat. Rev. Genet., 950–962.

51. Romanov, M.N., Farré, M., Lithgow, P.E., Fowler, K.E., Skinner, B.M., O'Connor, R., Fonseka, G., Backström, N., Matsuda, Y., Nishida, C., *et al.* (2014). Reconstruction of gross avian genome structure, organization and evolution suggests that the chicken lineage most closely resembles the dinosaur avian ancestor. BMC Genomics *15*, 1060.

52. Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S., and Miller, W. (2007).

Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. *17*, 413–421.

53. Richard, F., Lombard, M., and Dutrillaux, B. (2003). Reconstruction of the ancestral karyotype of eutherian mammals. Chromosom. Res. *11*, 605–618.

54. Beklemisheva, V.R., Perelman, P.L., Lemskaya, N.A., Kulemzina, A.I., Proskuryakova, A.A., Burkanov, V.N., and Graphodatsky, A.S. (2016). The ancestral carnivore karyotype as substantiated by comparative chromosome painting of three pinnipeds, the walrus, the steller sea lion and the baikal seal (Pinnipedia, Carnivora). PLoS One *11*.

55. De Leo, A.A., Guedelha, N., Toder, R., Voullaire, L., Ferguson-Smith, M.A., O'Brien, P.C.M., and Graves, J.A.M. (1999). Comparative chromosome painting between marsupial orders: Relationships with a 2n = 14 ancestral marsupial karyotype. Chromosom. Res. *7*, 509–517.

56. Stanyon, R., Rocchi, M., Capozzi, O., Roberto, R., Misceo, D., Ventura, M., Cardone, M.F., Bigoni, F., and Archidiacono, N. (2008). Primate chromosome evolution: Ancestral karyotypes, marker order and neocentromeres. Chromosom. Res. *16*, 17–39.

57. Sankoff, D. (2012). Edit distance for genome comparison based on non-local operations. In.

58. Ma, J., Zhang, L., Suh, B.B., Raney, B.J., Burhans, R.C., Kent, W.J., Blanchette, M., Haussler, D., and Miller, W. (2006). Reconstructing contiguous regions of an ancestral genome. Genome Res. *16*, 1557–1565.

59. Meidanis, J., Porto, O., and Telles, G.P. (1998). On the consecutive ones property. Discret. Appl. Math. *88*, 325–354.

60. Jones, B.R., Rajaraman, A., Tannier, E., and Chauve, C. (2012). ANGES:

Reconstructing ancestral genomes maps. Bioinformatics *28*, 2388–2390.

61. Perrin, A., Varré, J.S., Blanquart, S., and Ouangraoua, A. (2015). ProCARs: Progressive Reconstruction of Ancestral Gene Orders. BMC Genomics *16*.
62. Gagnon, Y., Blanchette, M., and El-Mabrouk, N. (2012). A flexible ancestral genome reconstruction method based on gapped adjacencies. BMC Bioinformatics *13*.
63. Bertrand, D., Gagnon, Y., Blanchette, M., and El-Mabrouk, N. (2010). Reconstruction of ancestral genome subject to whole genome duplication, speciation, rearrangement and loss. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 78–89.
64. SANKOFF, D., and BLANCHETTE, M. (1998). Multiple Genome Rearrangement and Breakpoint Phylogeny. J. Comput. Biol. *5*, 555–570.
65. Mcgeoch, C.C. (1996). Toward an experimental method for algorithm simulation. INFORMS J. Comput.
66. Moret, B.M.E., Bader, D.A., Wyman, S., Warnow, T., and Yan, M. (2013). A New Implementation and Detailed Study of Breakpoint Analysis. In Pacific Symposium on Biocomputing, pp. 583–594.
67. Bourque, G., and Pevzner, P.A. (2002). Genome-scale evolution: Reconstructing gene orders in the ancestral species. Genome Res. *12*, 26–36.
68. Alekseyev, M.A., and Pevzner, P.A. (2009). Breakpoint graphs and ancestral genome reconstructions. Genome Res. *19*, 943–957.
69. Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D.M., Ma, J., and Lewin, H.A. (2017). Reconstruction and evolutionary history of eutherian chromosomes. Proc. Natl. Acad. Sci. *114*, 5379–5388.

70. Damas, J., Kim, J., Farré, M., Griffin, D.K., and Larkin, D.M. (2018). Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. Genome Biol., 155.

71. Farré, M., Kim, J., Proskuryakova, A.A., Zhang, Y., Kulemzina, A.I., Li, Q., Zhou, Y., Xiong, Y., Johnson, J.L., Perelman, P.L., *et al.* (2019). Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. Genome Res. *29*, 576–589.

72. Suckow, M.A., Danneman, P., and Brayton, C. (2000). The laboratory mouse.

73. IUCN (International Union for Conservation and Nature) (2019). The IUCN Red List of Threatened Species. Version 2019-1.

74. Burgin, C.J., Colella, J.P., Kahn, P.L., and Upham, N.S. (2018). How many species of mammals are there? J. Mammal. *99*, 1–14.

75. Liu, Y., Hu, W., Wang, H., Lu, M., Shao, C., Menzel, C., Yan, Z., Li, Y., Zhao, S., Khaitovich, P., *et al.* (2010). Genomic analysis of miRNAs in an extreme mammalian hibernator, the Arctic ground squirrel. Physiol. Genomics *42*, 39–51.

76. Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A. V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., *et al.* (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223–227.

77. Fang, X., Seim, I., Huang, Z., Gerashchenko, M. V., Xiong, Z., Turanov, A.A., Zhu, Y., Lobanov, A. V., Fan, D., Yim, S.H., *et al.* (2014). Adaptations to a Subterranean Environment and Longevity Revealed by the Analysis of Mole Rat Genomes. Cell Rep. *8*, 1354–1364.

78. Buffenstein, R. (2005). The naked mole-rat: A new long-living model for human

aging research. Journals Gerontol. - Ser. A Biol. Sci. Med. Sci. *60*, 1369–1377.

79. Seluanov, A., Hine, C., Azpurua, J., Feigenson, M., Bozzella, M., Mao, Z., Catania, K.C., and Gorbunova, V. (2009). Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. Proc. Natl. Acad. Sci. U. S. A. *106*, 19352–19357.

80. Keane, M., Craig, T., Alföldi, J., Berlin, A.M., Johnson, J., Seluanov, A., Gorbunova, V., Di Palma, F., Lindblad-Toh, K., Church, G.M., *et al.* (2014). The Naked Mole Rat Genome Resource: Facilitating analyses of cancer and longevity-related adaptations. Bioinformatics *30*, 3558–3560.

81. Park, T.J., Reznick, J., Peterson, B.L., Blass, G., Omerbašić, D., Bennett, N.C., Kuich, P.H.J.L., Zasada, C., Browe, B.M., Hamann, W., *et al.* (2017). Fructose-driven glycolysis supports anoxia resistance in the naked mole-rat. Science (80-.). *356*, 307–311.

82. Park, T.J., Lu, Y., Jüttner, R., Smith, E.S.J., Hu, J., Brand, A., Wetzel, C., Milenkovic, N., Erdmann, B., Heppenstall, P.A., *et al.* (2008). Selective inflammatory pain insensitivity in the African naked mole-rat (Heterocephalus glaber). PLoS Biol. *6*.

83. Huchon, D., Madsen, O., Sibbald, M.J.J.B., Ament, K., Stanhope, M.J., Catzeflis, F., De Jong, W.W., and Douzery, E.J.P. (2002). Rodent phylogeny and a timescale for the evolution of glires: Evidence from an extensive taxon sampling using three nuclear genes. Mol. Biol. Evol. *19*, 1053–1065.

84. Blanga-Kanfi, S., Miranda, H., Penn, O., Pupko, T., Debry, R.W., and Huchon, D. (2009). Rodent phylogeny revised: Analysis of six nuclear genes from all major rodent clades. BMC Evol. Biol. *9*.

85. Churakov, G., Sadasivuni, M.K., Rosenbloom, K.R., Huchon, D., Brosius, J.,

and Schmitz, J. (2010). Rodent evolution: Back to the root. Mol. Biol. Evol. *27*, 1315–1326.

86. Graur, D., Hide, W.A., and Li, W.H. (1991). Is the guinea-pig a rodent? Nature *351*, 649–652.

87. D'erchia, A.M., Gissi, C., Pesole, G., Saccone, C., and Arnason, U. (1996). The guinea-pig is not a rodent. Nature *381*, 597–600.

88. Luckett, W.P., and Hartenberger, J.L. (1993). Monophyly or polyphyly of the order Rodentia: Possible conflict between morphological and molecular interpretations. J. Mamm. Evol. *1*, 127–147.

89. Frye, M.S., and Hedges, S.B. (1995). Monophyly of the order rodentia inferred from mitochondrial DNA sequences of the genes for 12S rRNA, 16S rRNA, and tRNA-valine. Mol. Biol. Evol. *12*, 168–176.

90. Huchon, D., Catzeflis, F.M., and Douzery, E.J.P. (2000). Variance of molecular datings, evolution of rodents and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. Proc. R. Soc. B Biol. Sci. *267*, 393–402.

91. Silva, M.J., and Yonenaga-Yassuda, Y. (1998). Karyotype and chromosomal polymorphism of an undescribed Akodon from Central Brazil, a species with the lowest known diploid chromosome number in rodents. Cytogenet. Cell Genet. *81*, 46–50.

92. Contreras, L.C., Torres-Mura, J.C., and Spotorno, A.E. (1990). The largest known chromosome number for a mammal, in a South American desert rodent. Experientia *46*, 506–508.

93. Gallardo, M.H., Bickham, J.W., Honeycutt, R.L., Ojeda, R.A., and Köhler, N. (1999). Discovery of tetraploidy in a mammal. Nature *401*, 341.

94. Gallardo, M.H., González, C.A., and Cebrián, I. (2006). Molecular cytogenetics

and allotetraploidy in the red vizcacha rat, Tympanoctomys barrerae (Rodentia, Octodontidae). Genomics *88*, 214–221.

95. Gallardo, M.H., Kausel, G., Jiménez, A., Bacquet, C., González, C., Figueroa, J., Köhler, N., and Ojeda, R. (2004). Whole-genome duplications in South American desert rodents (Octodontidae). In Biological Journal of the Linnean Society, pp. 443–451.

96. Svartman, M., Stone, G., and Stanyon, R. (2005). Molecular cytogenetics discards polyploidy in mammals. Genomics *85*, 425–430.

97. Evans, B.J., Upham, N.S., Golding, G.B., Ojeda, R.A., and Ojeda, A.A. (2017). Evolution of the largest mammalian genome. Genome Biol. Evol. *6*, 1711–1724.

98. Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol. Evol. *34*, 1812–1819.

99. Fagundes, V., Christoff, A.U., and Yonenaga-Yassuda, Y. (2004). Extraordinary Chromosomal Polymorphism with 28 different Karyotypes in the Neotropical Species Akodon Cursor (Muridae, Sigmodontinae), one of the Smallest Diploid Number in Rodents (2n = 16, 15 and 14). Hereditas *129*, 263–274.

100. Fagundes, V., Vianna-Morgante, A.M., and Yonenaga-Yassuda, Y. (1997). Telomeric sequences localization and G-banding patterns in the identification of a polymorphic chromosomal rearrangement in the rodent Akodon cursor (2n = 14, 15 and 16). Chromosom. Res. *5*, 228–232.

101. Church, D.M., Goodstadt, L., Hillier, L.W., Zody, M.C., Goldstein, S., She, X., Bult, C.J., Agarwala, R., Cherry, J.L., DiCuccio, M., *et al.* (2009). Lineage-specific biology revealed by a finished genome assembly of the mouse. PLoS Biol. *5*.

102. Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.C., Agarwala, R., McLaren, W.M., Ritchie, G.R.S., *et al.* (2011). Modernizing reference genome assemblies. PLoS Biol. *7*.

103. Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., *et al.* (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature.

104. Lowe, T. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

105. Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., *et al.* (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol.

106. Fang, X., Nevo, E., Han, L., Levanon, E.Y., Zhao, J., Avivi, A., Larkin, D., Jiang, X., Feranchuk, S., Zhu, Y., *et al.* (2014). Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. Nat. Commun., 3966.

107. Di Palma, F., Alfoldi, J., Johnson, J., Berlin, A., Gnerre, S., Jaffe, D., MacCallum, I., Young, S., Walker, B.J., and Liblad-Toh, K. (2012). The Draft Genome of Jaculus jaculus.

108. Gnerre, S., Heiman, D., Young, S., Muzny, D.M., Nazareth, L., Weinstock, G., and Gibbs, R. (2014). Dipodomys ordii isolate 6190, whole genome shotgun sequence project.

109. Di Palma, F., Alfoldi, J., Johnson, J., Berlin, A., Gnerre, S., Jaffe, D., MacCallum, I., Young, S., Walker, B.J., and Lindblad-Toh, K. (2013). The Draft Genome of Spermophilus tridecemlineatus.

110. Culibrk, L., Leelakumari, S., Taylor, G.A., Tse, K., Cheung, D., Chuah, E., Kirk,

H., Pandoh, P., Troussard, A., Zhao, Y., *et al.* (2018). Yellow bellied marmot (Marmota flaviventris) genome sequencing and assembly.

111. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., *et al.* (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.

112. Tomasco, I.H., and Lessa, E.P. (2011). The evolution of mitochondrial genomes in subterranean caviomorph rodents: Adaptation against a background of purifying selection. Mol. Phylogenet. Evol. *61*, 64–70.

113. Collins, F.S., Lander, E.S., Rogers, J., and Waterson, R.H. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.

114. Gissi, C., Gullberg, A., and Arnason, U. (1998). The complete mitochondrial DNA sequence of the rabbit, Oryctolagus cuniculus. Genomics *50*, 161–169.

115. Lok, S., Paton, T.A., Wang, Z., Kaur, G., Walker, S., Yuen, R.K.C., Sung, W.W.L., Whitney, J., Buchanan, J.A., Trost, B., *et al.* (2017). De Novo Genome and Transcriptome Assembly of the Canadian Beaver (Castor canadensis). G3 Genes|Genomes|Genetics *7*, 755–773.

116. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., *et al.* (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science (80-.). *356*, 92–95.

117. Dudchenko, O., Shamim, M.S., Batra, S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., Hilaire, B.G.S., Yao, W., Stamenova, E., *et al.* (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv.

118. Di Palma, F., Alfoldi, J., Johnson, J., Berlin, A., Gnerre, S., Jaffe, D., MacCallum, I., Young, S., Walker, B.J., and Lindblad-Toh, K. (2013). The Draft Genome of Chinchilla lanigera Available at: https://www.ncbi.nlm.nih.gov/nuccore/AGCD00000000.1.

119. Rambaut, A. (2009). FigTree, a graphical viewer of phylogenetic trees. Inst. Evol. Biol. Univ. Edinburgh.

120. Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. Pacific Symp. Biocomput. *7*, 115–126.

121. Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res. *13*, 103–107.

122. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: Enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

123. Kuhn, R.M., Haussler, D., and James Kent, W. (2013). The UCSC genome browser and associated tools. Brief. Bioinform. *14*, 144–161.

124. Harris, R.S. (2007). Improved Pairwise Alignment of Genomic DNA. Pennsylvania State Univ. PhD Thesis.

125. Romanenko, S.A., Volobouev, V.T., Perelman, P.L., Lebedev, V.S., Serdukova, N.A., Trifonov, V.A., Biltueva, L.S., Nie, W., O'Brien, P.C.M., Bulatova, N.S., *et al.* (2007). Karyotype evolution and phylogenetic relationships of hamsters (Cricetidae, Muroidea, Rodentia) inferred from chromosomal painting and banding comparison. Chromosom. Res. *15*, 283–298.

126. Romanenko, S.A., Perelman, P.L., Trifonov, V.A., and Graphodatsky, A.S. (2012). Chromosomal evolution in Rodentia. Heredity (Edinb)., 4–16.

127. Engelbrecht, A., Dobigny, G., and Robinson, T.J. (2006). Further insights into the ancestral murine karyotype: The contribution of the Otomys-Mus comparison using chromosome painting. Cytogenet. Genome Res. *112*, 126–130.

128. Romanenko, S.A., Perelman, P.L., Serdukova, N.A., Trifonov, V.A., Biltueva, L.S., Wang, J., Li, T., Nie, W., O'Brien, P.C.M., Volobouev, V.T., *et al.* (2006). Reciprocal chromosome painting between three laboratory rodent species. Mamm. Genome *17*, 1183–1192.

129. Tesler, G. (2002). GRIMM: Genome rearrangements web server. Bioinformatics *18*, 492–493.

130. Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in mammalian tissues. Science (80-.). *338*, 1593–1599.

131. Bens, M., Szafranski, K., Holtze, S., Sahm, A., Groth, M., Kestler, H.A., Hildebrandt, T.B., and Platzer, M. (2018). Naked mole-rat transcriptome signatures of socially suppressed sexual maturation and links of reproduction to aging. BMC Biol. *16*, 77.

132. Schmidt, H., Malik, A., Bicker, A., Poetzsch, G., Avivi, A., Shams, I., and Hankeln, T. (2017). Hypoxia tolerance, longevity and cancer-resistance in the mole rat Spalax - A liver transcriptomics approach. Sci. Rep. *7*.

133. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

134. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

135. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

136. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*.

137. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440.

138. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

139. Quinlan, A.R. (2014). BEDTools: The Swiss-Army tool for genome feature analysis. Curr. Protoc. Bioinforma. *47*.

140. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit. Anal., 199–236.

141. Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47*, 419–426.

142. Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. *14*, 703–721.

143. Hsu, T.C., and Benirschke, K. (2006). Atlas of Mammalian Chromosomes S. J. O'Brien, J. C. Menninger, and W. G. Nash, eds. (John Wiley & Sons Inc).

144. Shahin, A.A.B., and Ata, A.T.M. (2001). A comparative study on the karyotype and meiosis of the jerboas allactaga and jaculus (rodentia: dipodidae) in egypt.

Zool. Middle East *22*, 5–16.

145. Bostock, C.J., Christie, S., Hatch, F.T., and Mazrimas, J.A. (1977). A stable, “near-haploid” mammalian cell line (Dipodomys ordii). Exp. Cell Res. *106*, 373–377.

146. Nadler, C.F., and Hughes, C.E. (1966). Chromosomes and Taxonomy of the Ground Squirrel Subgenus Ictidomys. J. Mammal. *47*, 46–53.

147. Hoffmann, R.S., and Nadler, C.F. (1968). Chromosomes and systematics of some north American species of the genus Marmota (Rodentia: Sciuridae). Experientia *24*, 740–742.

148. Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Sanchez, R.C., Fescemyer, H.W., Harris, R., Ye, D., O’Brien, P.C.M., Chikhi, R., Ryder, O.A., *et al.* (2016). A time- and cost-effective strategy to sequence mammalian Y chromosomes: An application to the de novo assembly of gorilla Y. Genome Res. *26*, 530–540.

149. Fabre, P.H., Hautier, L., Dimitrov, D., and P Douzery, E.J. (2012). A glimpse on the pattern of rodent diversification: A phylogenetic approach. BMC Evol. Biol. *12*.

150. Montgelard, C., Forty, E., Arnal, V., and Matthee, C.A. (2008). Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evol. Biol. *8*.

151. Lee, J., Lee, D., Sim, M., Kwon, D., Kim, J., Ko, Y., and Kim, J. (2018). mySyntenyPortal: An application package to construct websites for synteny block analysis. BMC Bioinformatics *19*.

152. Froenicke, L., Caldés, M.G., Graphodatsky, A., Müller, S., Lyons, L.A., Robinson, T.J., Volleth, M., Yang, F., and Wienberg, J. (2006). Are molecular

cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? Genome Res., 306–310.

153. Li, T., O'Brien, P.C.M., Biltueva, L., Fu, B., Wang, J., Nie, W., Ferguson-Smith, M.A., Graphodatsky, A.S., and Yang, F. (2004). Evolution of genome organizations of squirrels (Sciuridae) revealed by cross-species chromosome painting. Chromosom. Res. *12*, 317–335.

154. Richard, F., Messaoudi, C., Bonnet-Garnier, A., Lombard, M., and Dutrillaux, B. (2003). Highly conserved chromosomes in an Asian squirrel (Menetes berdmorei, Rodentia: Sciuridae) as demonstrated by ZOO-FISH with human probes. Chromosom. Res. *11*, 597–603.

155. Stanyon, R., Stone, G., Garcia, M., and Froenicke, L. (2003). Reciprocal chromosome painting shows that squirrels, unlike murid rodents, have a highly conserved genome organization. Genomics *82*, 245–249.

156. Li, T., Wang, J., Su, W., Nie, W., and Yang, F. (2006). Karyotypic evolution of the family Sciuridae: Inferences from the genome organizations of ground squirrels. Cytogenet. Genome Res. *112*, 270–276.

157. Farré, M., Micheletti, D., and Ruiz-Herrera, A. (2013). Recombination rates and genomic shuffling in human and chimpanzee - A new twist in the chromosomal speciation theory. Mol. Biol. Evol. *30*, 853–864.

158. Coyne, J.A., Aulard, S., and Berry, A. (1991). Lack of underdominance in a naturally occurring pericentric inversion in drosophila melanogaster and its implications for chromosome evolution. Genetics *129*, 791–802.

159. Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M.T.P., *et al.* (2018). Earth BioGenome Project: Sequencing life for the future of life. Proc. Natl. Acad. Sci.

115, 4325–4333.

160. Burt, D.W., Bruley, C., Dunn, I.C., Jones, C.T., Ramage, A., Law, A.S., Morrice, D.R., Paton, I.R., Smith, J., Windsor, D., *et al.* (1999). The dynamics of chromosome evolution in birds and mammals. Nature *402*, 411–413.

161. Bourque, G., Zdobnov, E.M., Bork, P., Pevzner, P.A., and Tesler, G. (2005). Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. Genome Res. *15*, 98–110.

162. Sudmant, P.H., Alexis, M.S., and Burge, C.B. (2015). Meta-analysis of RNA-seq expression data across species, tissues and studies. Genome Biol. *16*.

163. Lamanna, F., Kirschbaum, F., Waurick, I., Dieterich, C., and Tiedemann, R. (2015). Cross-tissue and cross-species analysis of gene expression in skeletal muscle and electric organ of African weakly-electric fish (Teleostei; Mormyridae). BMC Genomics *16*, 668.

164. Corbett-Detig, R.B., and Hartl, D.L. (2012). Population Genomics of Inversion Polymorphisms in Drosophila melanogaster. PLoS Genet. *8*.

165. Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., *et al.* (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat. Cell Biol., 305–310.

166. Feuk, L., MacDonald, J.R., Tang, T., Carson, A.R., Li, M., Rao, G., Khaja, R., and Scherer, S.W. (2005). Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet. *1*, 56.

167. Acemel, R.D., Maeso, I., and Gómez-Skarmeta, J.L. (2017). Topologically

associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip. Rev. Dev. Biol.

168. Krefting, J., Andrade-Navarro, M.A., and Ibn-Salem, J. (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. BMC Biol. *16*.

169. Lazar, N.H., Nevonen, K.A., O'Connell, B., McCann, C., O'Neill, R.J., Green, R.E., Meyer, T.J., Okhovat, M., and Carbone, L. (2018). Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. Genome Res. *28*, 983–997.

170. Larkin, D.M., Pape, G., Donthu, R., Auvil, L., Welge, M., and Lewin, H.A. (2009). Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res. *19*, 770–777.

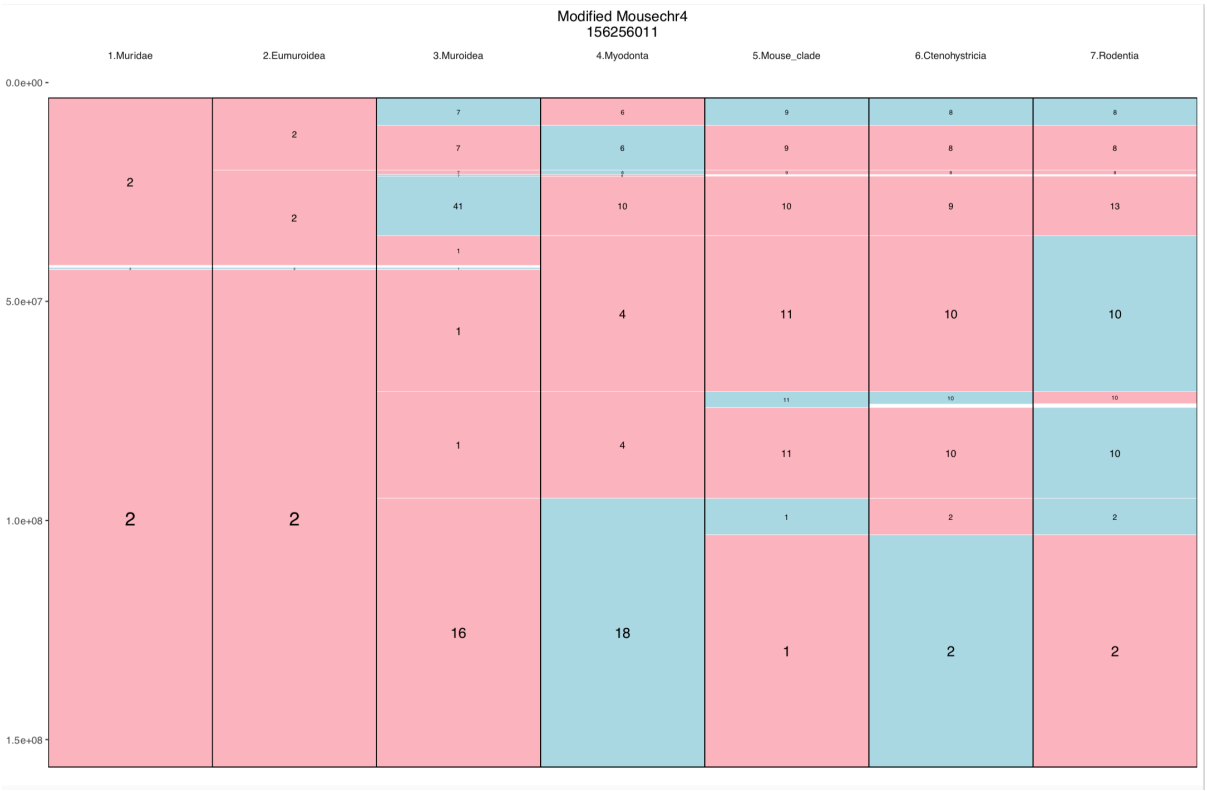
171. Ullastres, A., Farré, M., Capilla, L., and Ruiz-Herrera, A. (2014). Unraveling the effect of genomic structural changes in the rhesus macaque - implications for the adaptive role of inversions. BMC Genomics *15*, 530.

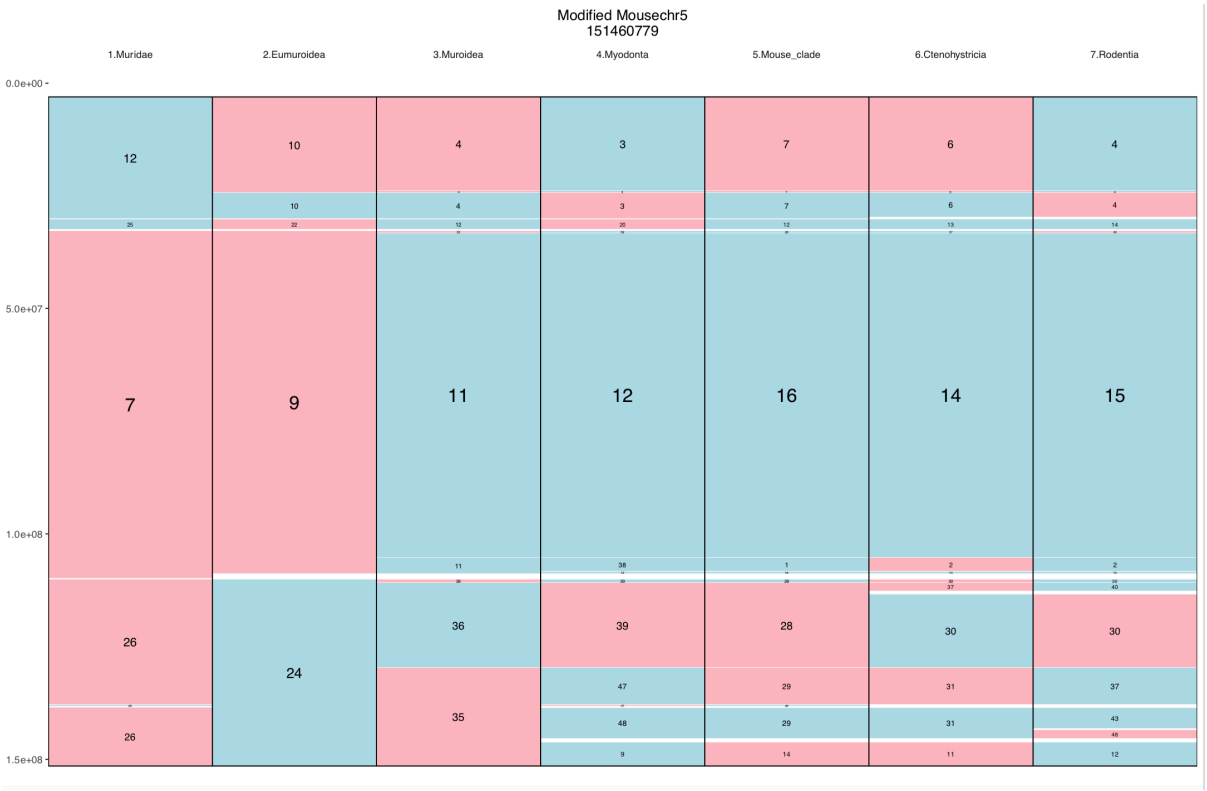
Appendix


Modified Mousechr1
195225877
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08
2.0e+08

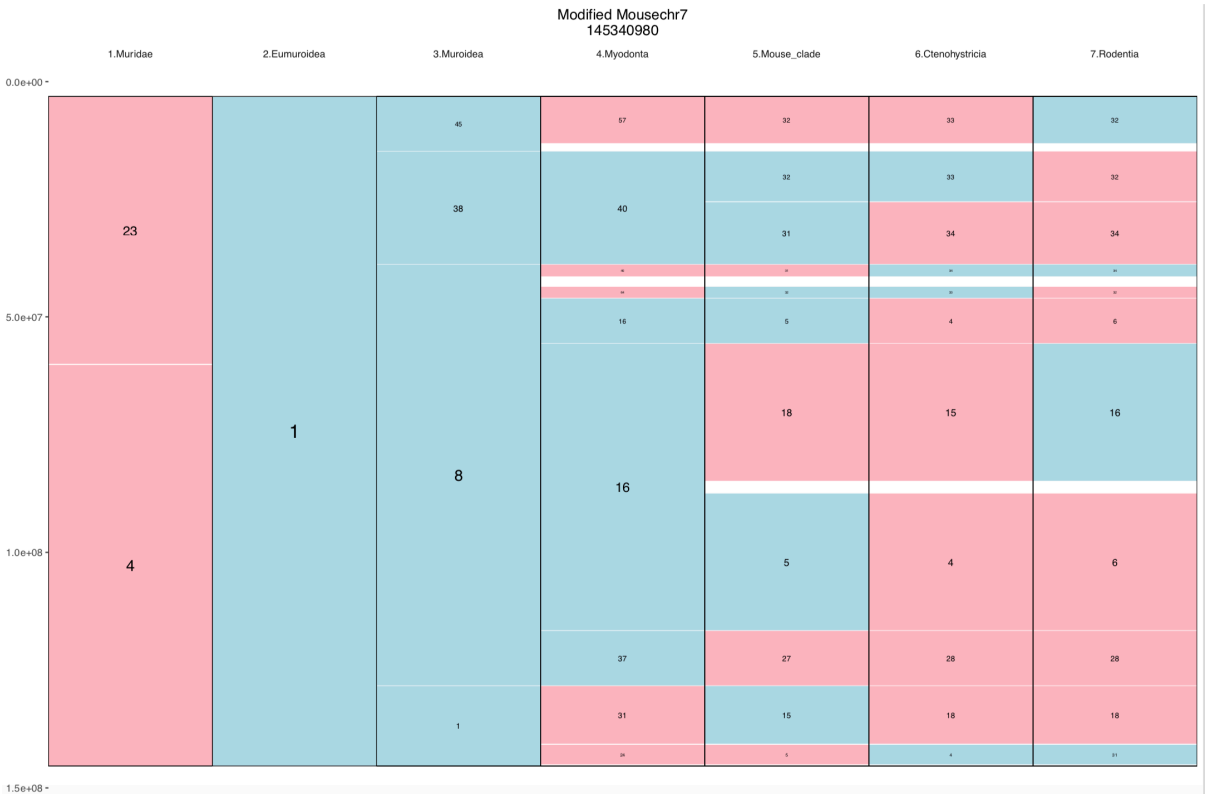

Modified Mousechr2
181884607
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08


Modified Mousechr3
159939176
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08

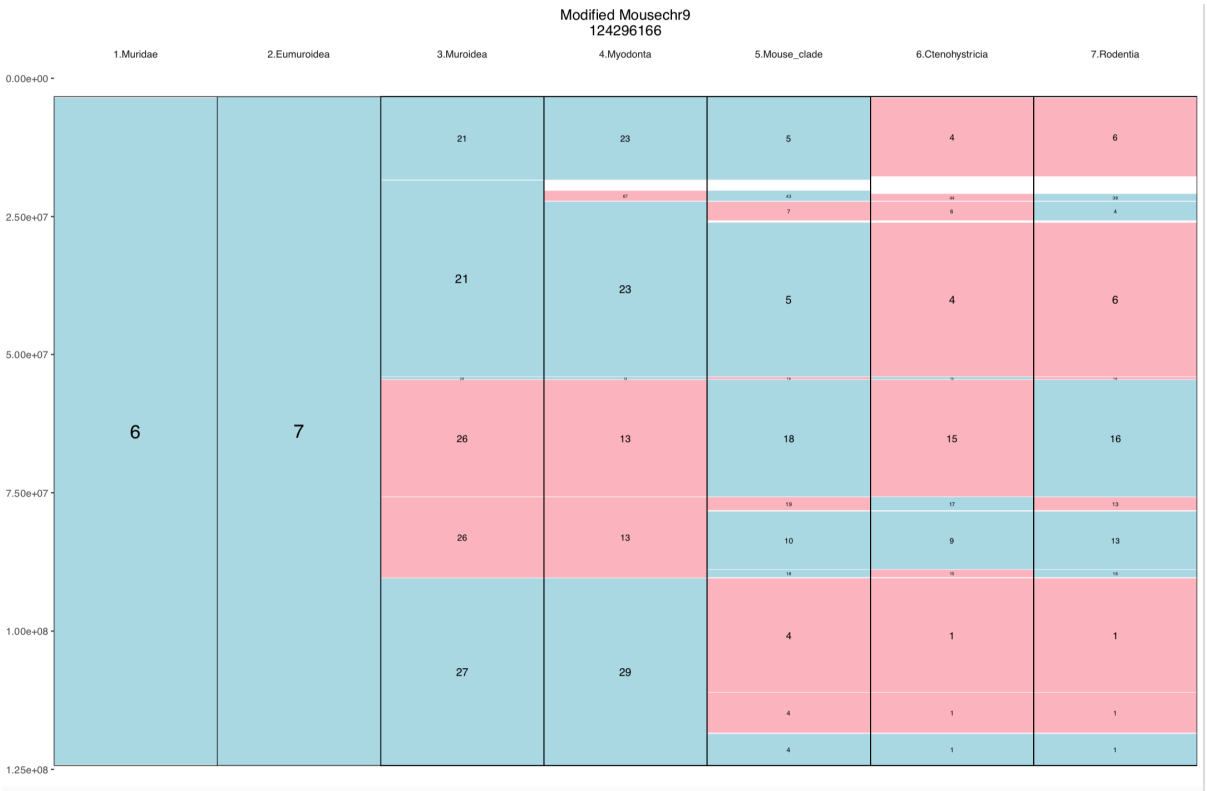

Modified Mousechr4
156256011
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08

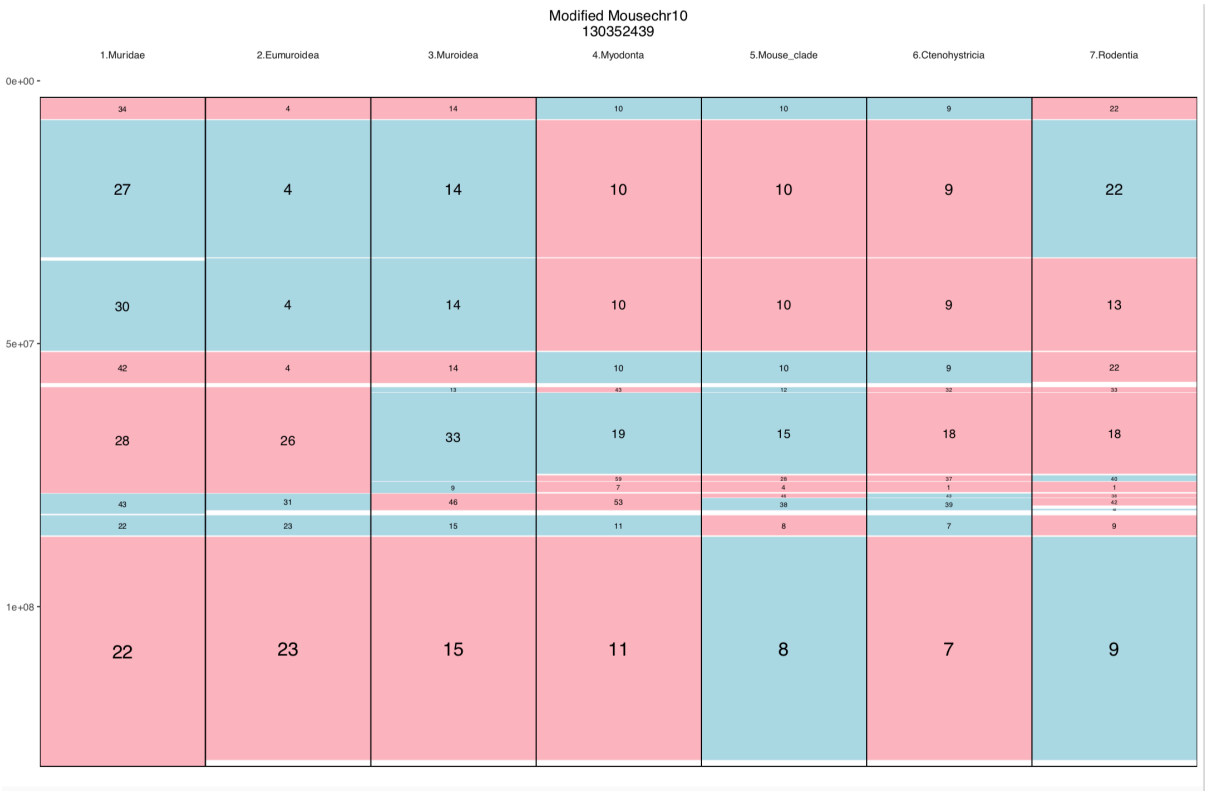

Modified Mousechr5
151460779
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08


Modified Mousechr6
149586546
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08

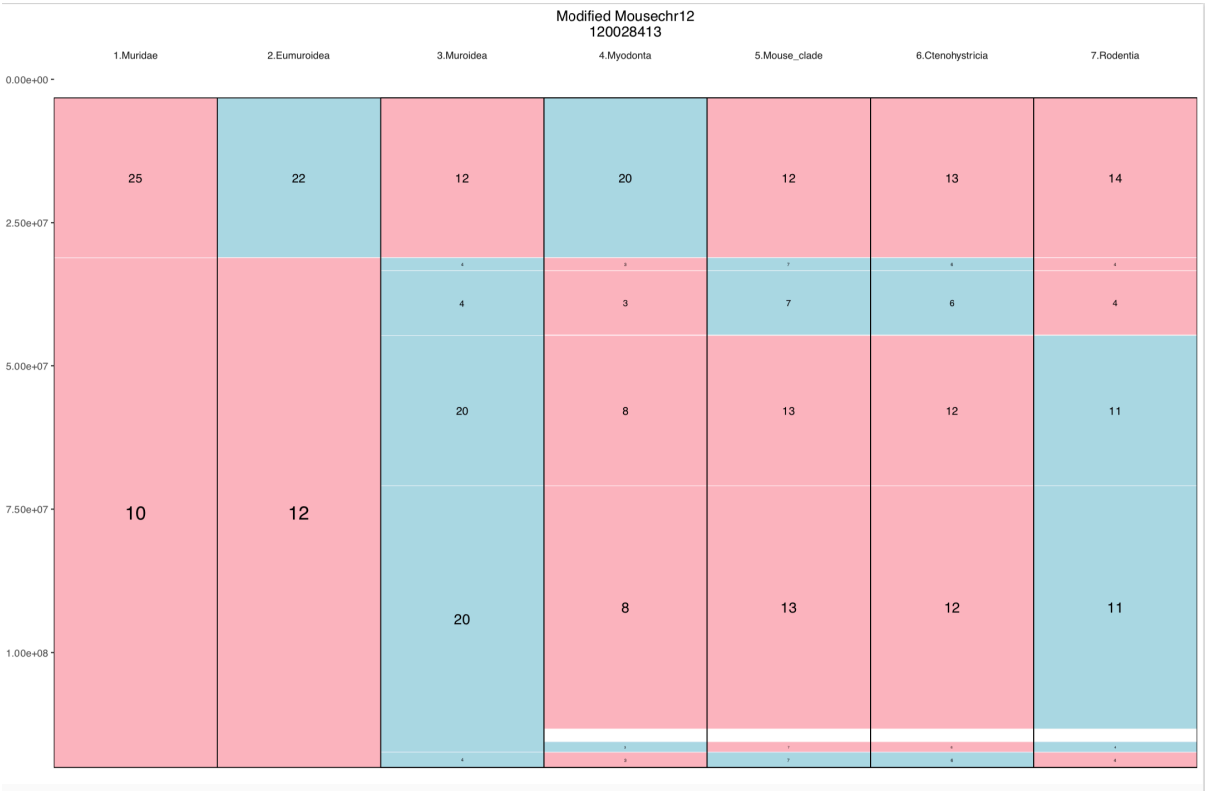

Modified Mousechr7
145340980
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
5.0e+07
1.0e+08
1.5e+08


Modified Mousechr8
129300768
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0e+00
5e+07
1e+08

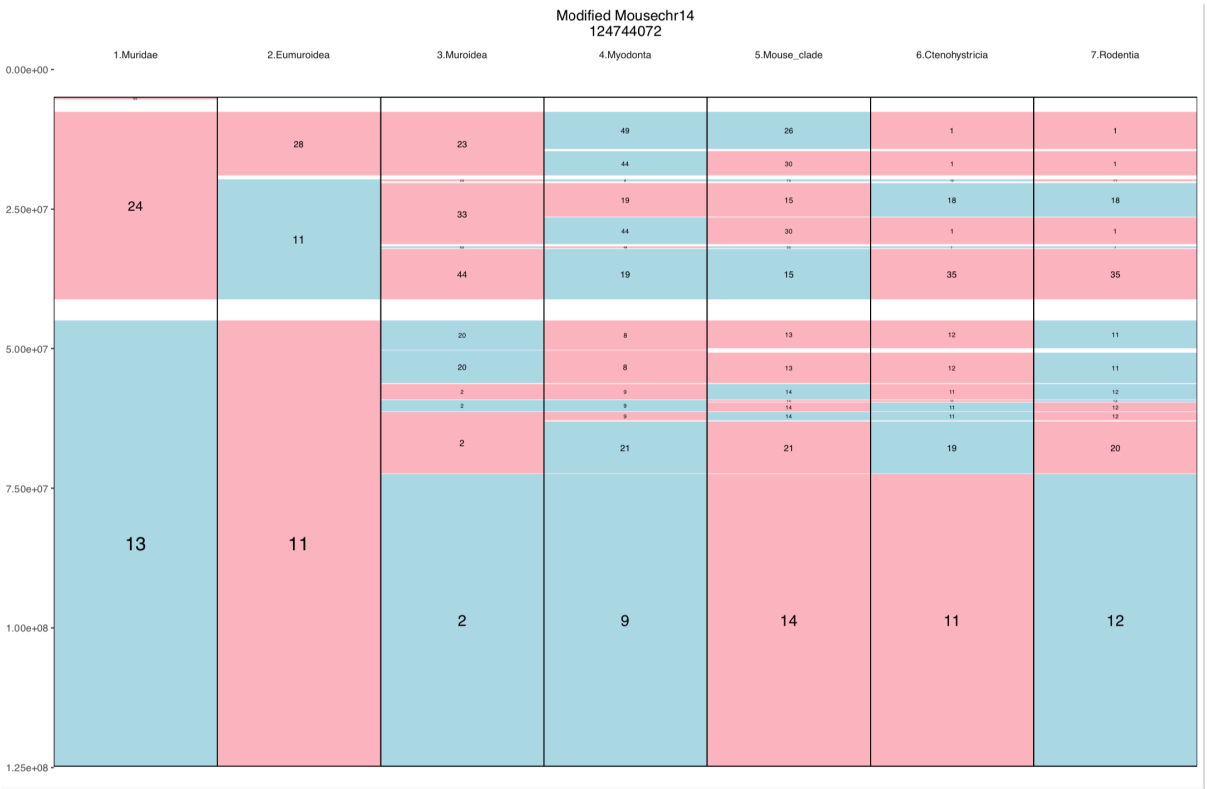

Modified Mousechr9
124296166
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.00e+00
2.50e+07
5.00e+07
7.50e+07
1.00e+08
1.25e+08


Modified Mousechr10
130352439
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0e+00
5e+07
1e+08


Modified Mousechr11
121874811
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.00e+00
2.50e+07
5.00e+07
7.50e+07
1.00e+08
1.25e+08
7
9
12
46
33
36
4
20
12
13
14
3
3
32
2
3
3
8
8
5
17
22
24
24
29
29

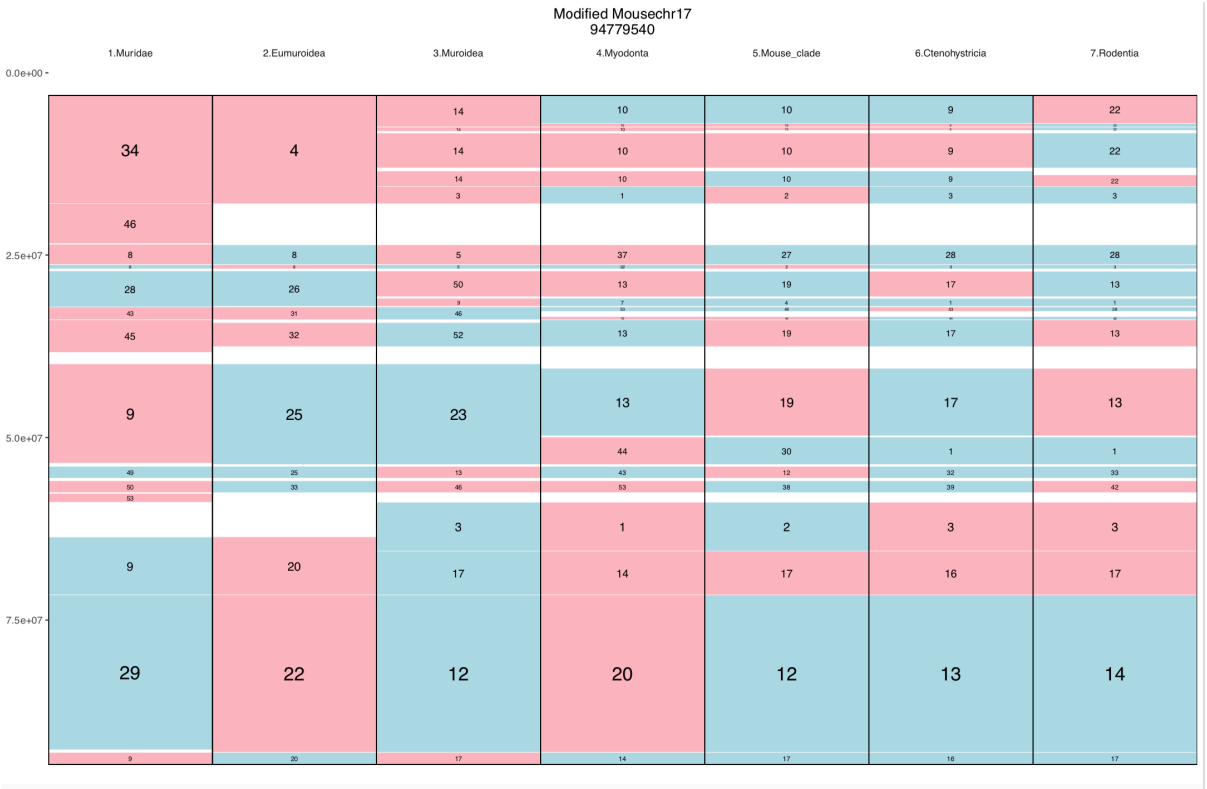

Modified Mousechr12
120028413
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.00e+00
2.50e+07
5.00e+07
7.50e+07
1.00e+08
25
22
12
20
12
13
14
4
3
7
6
4
20
8
13
12
11
10
12
20
8
13
12
11


Modified Mousechr13
120320772
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.00e+00
2.50e+07
5.00e+07
7.50e+07
1.00e+08
17
16
32
28
28
40
34
30
30
1
54
25
7
19
19
39
39
2
36
26
6
17
17
40
40
1
3
38
26
4
27
27
27
44
45
3
49
44
18
4
3
1
2
3
3
18
4
3
1
2
3
3


Modified Mousechr14
124744072
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.00e+00
2.50e+07
5.00e+07
7.50e+07
1.00e+08
1.25e+08
24
11
13
11
28
23
33
44
20
20
2
2
2
49
44
19
44
19
8
8
9
9
9
21
26
30
15
30
15
13
13
14
14
14
21
1
1
18
1
35
12
12
11
11
12
12
19
1
1
18
1
35
11
11
12
12
20
2
9
14
11
12


Modified Mousechr15
103943617
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
2.5e+07
5.0e+07
7.5e+07
1.0e+08
18
4
3
1
2
3
3
19
17
7
6
9
8
8


Modified Mousechr16
98011824
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
2.5e+07
5.0e+07
7.5e+07
1.0e+08
8
8
5
37
27
28
28
14
14
9
7
4
1
1


Modified Mousechr17
94779540
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
2.5e+07
5.0e+07
7.5e+07

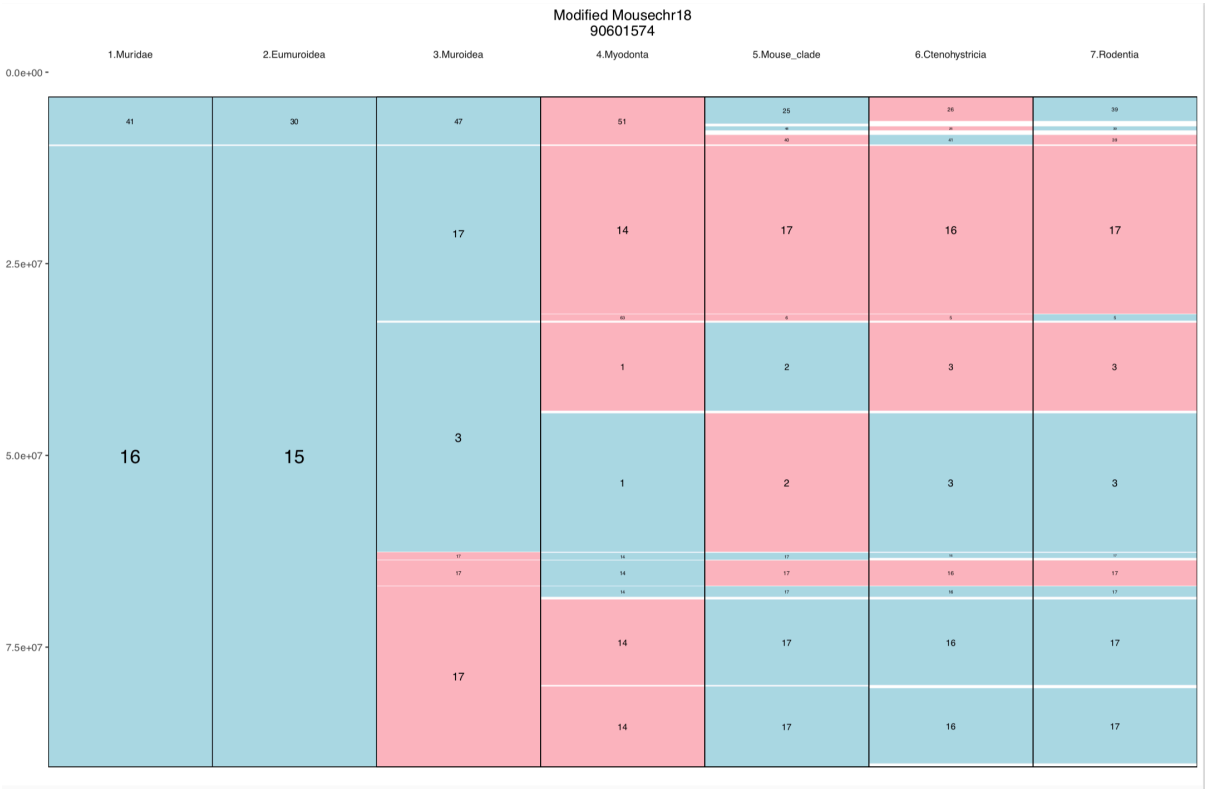

Modified Mousechr18
90601574
1.Muridae
2.Eumuroidea
3.Muroidea
4.Myodonta
5.Mouse_clade
6.Ctenohystricia
7.Rodentia
0.0e+00
2.5e+07
5.0e+07
7.5e+07



Figure 32 - *Mus musculus* chromosomes with respect to reconstructed ancestors after manual merging of APCFs

Table 8 - Gene counts for RNA-Seq data

Reference	Species	Tissue	No Feature	Ambiguous	Not Unique	Total	Unique
SRR594397_1	*Mus musculus*	Liver	9,268,845	4,446,198	16,997,857	99,731,984	69,019,084
SRR594397_2	*Mus musculus*	Liver	9,214,586	4,234,414	18,390,368	99,686,921	67,847,553
SRR594401_1	*Mus musculus*	Testes	16,506,139	3,781,325	8,876,354	106,009,704	76,845,886
SRR594401_2	*Mus musculus*	Testes	16,503,981	3,887,597	8,737,421	104,844,631	75,715,632
SRR594405_1	*Mus musculus*	Liver	10,891,001	6,425,387	11,605,791	124,699,470	95,777,291
SRR594405_2	*Mus musculus*	Liver	10,569,811	6,326,421	12,475,158	122,022,322	92,650,932
SRR594409_1	*Mus musculus*	Testes	16,753,902	4,814,917	6,568,999	112,711,272	84,573,454
SRR594409_2	*Mus musculus*	Testes	16,437,187	4,950,727	6,211,145	110,351,794	82,752,735
SRR594414_1	*Mus musculus*	Liver	3,275,300	1,370,923	4,695,755	32,441,073	23,099,095
SRR594414_2	*Mus musculus*	Liver	3,132,128	1,333,204	4,585,940	30,987,310	21,936,038
SRR594418_1	*Mus musculus*	Testes	4,344,404	1,165,093	3,395,780	33,428,470	24,523,193
SRR594418_2	*Mus musculus*	Testes	4,374,775	1,212,391	3,345,569	33,432,845	24,500,110
SRR594423_1	*Rattus norvegicus*	Liver	3,566,676	199,984	3,555,100	24,673,117	17,351,357
SRR594423_2	*Rattus norvegicus*	Liver	3,473,747	178,563	3,528,180	24,632,583	17,452,093
SRR594427_1	*Rattus norvegicus*	Testes	12,043,023	1,059,635	27,587,307	110,893,902	70,203,937
SRR594427_2	*Rattus norvegicus*	Testes	12,213,208	990,892	26,714,642	109,476,163	69,557,421
SRR594432_1	*Rattus norvegicus*	Liver	17,317,213	1,304,717	12,718,066	126,715,587	95,375,591
SRR594432_2	*Rattus norvegicus*	Liver	16,705,750	1,284,430	12,201,709	125,192,761	95,000,872
SRR594436_1	*Rattus norvegicus*	Testes	24,390,668	1,064,574	11,171,194	109,878,318	73,251,882
SRR594436_2	*Rattus norvegicus*	Testes	24,385,059	1,005,076	11,077,998	109,298,572	72,830,439
SRR594441_1	*Rattus norvegicus*	Liver	7,256,259	341,515	5,947,549	39,580,645	26,035,322
SRR594441_2	*Rattus norvegicus*	Liver	6,841,385	313,463	5,733,566	38,222,867	25,334,453
SRR594445_1	*Rattus norvegicus*	Testes	4,625,729	394,698	5,922,208	38,942,184	27,999,549
SRR594445_2	*Rattus norvegicus*	Testes	4,530,133	361,345	5,597,277	37,423,550	26,934,795
SRR975606_1	*Fukomys Damarensis*	Liver	11,990,751	442,059	1,615,675	35,474,402	21,425,917
SRR975606_2	*Fukomys Damarensis*	Liver	11,766,939	433,525	1,586,552	34,790,345	21,003,329
SRR975609_1	*Fukomys Damarensis*	Liver	10,771,060	498,336	1,357,899	32,626,927	19,999,632
SRR975609_2	*Fukomys Damarensis*	Liver	10,572,450	489,508	1,335,432	32,025,871	19,628,481
SRR975613_1	*Fukomys Damarensis*	Liver	13,308,518	396,007	1,371,071	37,784,637	22,709,041
SRR975613_2	*Fukomys Damarensis*	Liver	13,063,950	389,452	1,355,007	37,067,880	22,259,471
SRR975616_1	*Fukomys Damarensis*	Liver	10,340,247	578,603	1,345,882	35,981,764	23,717,032
SRR975616_2	*Fukomys Damarensis*	Liver	10,148,534	567,937	1,321,694	35,301,183	23,263,018
SRR975612_1	*Fukomys Damarensis*	Testes	12,395,567	191,630	793,268	30,154,033	16,773,568
SRR975612_2	*Fukomys Damarensis*	Testes	12,174,344	187,943	779,038	29,611,644	16,470,319
SRR975617_1	*Fukomys Damarensis*	Testes	13,747,480	304,780	1,037,984	39,361,973	24,271,729
SRR975617_2	*Fukomys Damarensis*	Testes	13,577,877	300,699	1,029,628	38,874,717	23,966,513
SRR5516161	*Cavia porcellus*	Liver	5,942,106	547,165	1,397,903	23,517,576	15,630,402

SRR5516162	*Cavia porcellus*	Liver	6,560,466	568,910	1,491,933	27,955,736	19,334,427
SRR5516163	*Cavia porcellus*	Liver	5,831,424	474,116	1,346,210	26,573,243	18,921,493
SRR5516164	*Cavia porcellus*	Liver	6,473,119	485,480	1,316,768	24,407,531	16,132,164
SRR5516165	*Cavia porcellus*	Liver	6,697,767	442,118	1,250,926	22,597,697	14,206,886
SRR5516166	*Cavia porcellus*	Liver	6,839,045	549,780	1,570,573	30,428,658	21,469,260
SRR5516167	*Cavia porcellus*	Liver	6,020,760	776,587	2,053,601	27,788,069	18,937,121
SRR5516168	*Cavia porcellus*	Liver	5,231,033	485,282	1,247,349	23,311,949	16,348,285
SRR5516169	*Cavia porcellus*	Liver	5,482,018	461,175	1,365,087	24,334,643	17,026,363
SRR5516170	*Cavia porcellus*	Liver	5,359,746	652,981	1,490,866	26,132,439	18,628,846
SRR5516171	*Cavia porcellus*	Liver	5,661,588	664,298	1,637,210	27,045,537	19,082,441
SRR5516172	*Cavia porcellus*	Liver	6,393,992	554,517	1,512,089	27,818,081	19,357,483
SRR5516173	*Cavia porcellus*	Liver	5,662,488	479,412	1,318,211	22,681,737	15,221,626
SRR5516174	*Cavia porcellus*	Liver	5,630,821	482,502	1,251,321	23,236,396	15,871,752
SRR5516175	*Cavia porcellus*	Liver	8,103,614	589,217	1,651,217	30,162,627	19,818,579
SRR5516176	*Cavia porcellus*	Liver	7,091,556	663,828	1,712,808	29,622,793	20,154,601
SRR5516177	*Cavia porcellus*	Liver	9,852,091	634,913	1,679,517	31,739,544	19,573,023
SRR5516178	*Cavia porcellus*	Liver	7,576,641	692,174	1,715,412	31,844,707	21,860,480
SRR5516179	*Cavia porcellus*	Liver	6,046,165	542,884	1,387,690	25,325,572	17,348,833
SRR5516180	*Cavia porcellus*	Liver	6,792,231	576,486	1,453,004	26,478,654	17,656,933
SRR5516181	*Cavia porcellus*	Liver	6,625,889	482,092	1,387,038	25,050,227	16,555,208
SRR5516182	*Cavia porcellus*	Liver	6,868,964	663,671	1,656,287	30,054,323	20,865,401
SRR5516183	*Cavia porcellus*	Liver	5,538,475	469,515	1,414,382	25,286,118	17,863,746
SRR5516184	*Cavia porcellus*	Liver	5,032,521	626,644	1,442,225	26,720,687	19,619,297
SRR5516245	*Cavia porcellus*	Testes	7,473,644	507,961	1,271,686	27,956,317	18,703,026
SRR5516246	*Cavia porcellus*	Testes	7,350,856	509,841	1,487,509	26,002,457	16,654,251
SRR5516247	*Cavia porcellus*	Testes	8,149,829	520,097	1,342,368	28,819,860	18,807,566
SRR5516248	*Cavia porcellus*	Testes	5,664,351	394,841	1,001,173	20,344,258	13,283,893
SRR5516249	*Cavia porcellus*	Testes	5,651,508	357,280	953,920	20,318,067	13,355,359
SRR5516250	*Cavia porcellus*	Testes	5,720,131	390,790	975,236	21,865,659	14,779,502
SRR5516251	*Cavia porcellus*	Testes	8,887,370	549,621	1,456,158	31,212,561	20,319,412
SRR5516252	*Cavia porcellus*	Testes	7,103,257	469,334	1,176,680	25,479,132	16,729,861
SRR5516253	*Cavia porcellus*	Testes	6,899,321	455,993	1,132,545	25,016,941	16,529,082
SRR5516254	*Cavia porcellus*	Testes	7,038,750	512,768	1,156,576	26,952,231	18,244,137
SRR5516255	*Cavia porcellus*	Testes	7,083,795	514,765	1,153,258	27,044,264	18,292,446
SRR5516256	*Cavia porcellus*	Testes	7,396,939	483,887	1,220,650	26,864,263	17,762,787
SRR5517242	*Heterocephalus glaber*	Testes	6,501,000	237,090	1,600,310	20,242,762	11,904,362
SRR5517246	*Heterocephalus glaber*	Testes	7,548,914	322,949	2,404,014	28,357,649	18,081,772
SRR5517248	*Heterocephalus glaber*	Testes	6,191,167	312,600	2,328,348	26,231,121	17,399,006
SRR5517250	*Heterocephalus glaber*	Testes	6,771,626	326,546	2,277,094	26,849,783	17,474,517
SRR5517252	*Heterocephalus glaber*	Testes	7,701,895	362,750	2,709,805	31,352,829	20,578,379
SRR5517260	*Heterocephalus glaber*	Testes	6,889,695	271,842	1,980,721	24,010,545	14,868,287
SRR5517262	*Heterocephalus glaber*	Testes	6,018,628	277,691	2,111,585	23,800,044	15,392,140
SRR5517272	*Heterocephalus glaber*	Testes	8,435,574	344,094	2,673,642	31,268,559	19,815,249

SRR5517274	*Heterocephalus glaber*	Testes	8,345,777	416,909	2,989,117	34,377,109	22,625,306
SRR5517276	*Heterocephalus glaber*	Testes	6,640,063	312,798	2,321,914	26,423,401	17,148,626
SRR5517278	*Heterocephalus glaber*	Testes	6,273,302	258,575	1,985,827	23,298,903	14,781,199
SRR5517282	*Heterocephalus glaber*	Testes	7,525,514	334,126	2,581,947	29,752,081	19,310,494
SRR5517432	*Heterocephalus glaber*	Liver	4,006,002	425,722	1,417,026	22,385,500	16,536,750
SRR5517433	*Heterocephalus glaber*	Liver	4,716,783	473,102	1,744,997	27,330,234	20,395,352
SRR5517434	*Heterocephalus glaber*	Liver	5,666,680	420,550	1,554,147	24,121,215	16,479,838
SRR5517435	*Heterocephalus glaber*	Liver	5,730,577	390,403	1,395,514	24,652,509	17,136,015
SRR5517436	*Heterocephalus glaber*	Liver	4,237,845	426,304	1,510,188	22,250,566	16,076,229
SRR5517437	*Heterocephalus glaber*	Liver	4,840,212	394,372	1,380,247	21,522,431	14,907,600
SRR5517438	*Heterocephalus glaber*	Liver	5,178,739	527,641	1,688,548	27,527,455	20,132,527
SRR5517439	*Heterocephalus glaber*	Liver	3,856,265	451,153	1,455,584	24,344,086	18,581,084
SRR5517440	*Heterocephalus glaber*	Liver	4,668,208	441,822	1,532,930	21,353,141	14,710,181
SRR5517441	*Heterocephalus glaber*	Liver	3,987,473	523,179	1,391,876	24,040,741	18,138,213
SRR5517442	*Heterocephalus glaber*	Liver	5,680,311	713,253	1,782,203	31,383,421	23,207,654
SRR5517443	*Heterocephalus glaber*	Liver	4,148,577	500,090	1,520,813	24,557,065	18,387,585
SRR5517444	*Heterocephalus glaber*	Liver	4,733,698	702,286	1,754,526	29,696,599	22,506,089
SRR5517445	*Heterocephalus glaber*	Liver	4,745,374	397,913	1,603,414	24,245,609	17,498,908
SRR5517446	*Heterocephalus glaber*	Liver	6,623,666	716,972	2,275,383	35,637,786	26,021,765
SRR5517447	*Heterocephalus glaber*	Liver	5,223,237	462,157	1,525,215	23,830,924	16,620,315
SRR5517448	*Heterocephalus glaber*	Liver	5,686,898	774,113	1,984,198	31,108,796	22,663,587
SRR5517449	*Heterocephalus glaber*	Liver	3,562,371	515,782	1,380,797	23,300,486	17,841,536
SRR5517450	*Heterocephalus glaber*	Liver	5,343,361	522,450	1,661,178	27,724,516	20,197,527
SRR5517451	*Heterocephalus glaber*	Liver	5,879,464	528,094	1,871,548	28,107,587	19,828,481
SRR5517452	*Heterocephalus glaber*	Liver	3,955,434	453,825	1,364,901	22,398,365	16,624,205
SRR5517453	*Heterocephalus glaber*	Liver	3,801,977	582,642	1,483,294	25,200,487	19,332,574
SRR5517454	*Heterocephalus glaber*	Liver	5,510,265	506,396	1,740,926	28,438,944	20,681,357
ERR2587660	*Oryctolagus cuniculus*	Liver	4,606,543	79,812	944,523	19,270,323	13,639,445
ERR2587661	*Oryctolagus cuniculus*	Liver	5,221,122	71,149	952,157	22,885,533	16,641,105
ERR2587662	*Oryctolagus cuniculus*	Liver	4,756,617	116,241	1,236,380	24,924,615	18,815,377
ERR2587663	*Oryctolagus cuniculus*	Liver	3,958,416	76,840	887,088	17,415,636	12,493,292
ERR2587666	*Oryctolagus cuniculus*	Testes	6,380,723	133,074	1,018,939	21,639,218	14,106,482
ERR2587685	*Oryctolagus cuniculus*	Testes	11,666,259	113,688	1,809,028	29,094,033	15,505,058
ERR2587686	*Oryctolagus cuniculus*	Testes	2,713,935	31,700	694,997	10,133,095	6,692,463